

(1) Publication number:

**0 293 274** Δ1

**(2)** 

## **EUROPEAN PATENT APPLICATION**

(2) Application number: 88400790.7

2 Date of filing: 31.03.88

(a) Int. Cl.4: C 12 N 15/00 A 61 K 39/29

30 Priority: 31.03.87 JP 78313/87 04.06.87 JP 140586/87 10.11.87 JP 283990/87

- 43 Date of publication of application: 30.11.88 Bulletin 88/48
- Designated Contracting States:
  CH DE FR GB IT LI NL SE
- (7) Applicant: MITSUBISHI CHEMICAL INDUSTRIES LIMITED 5-2, Marunouchi 2-chome Chiyoda-ku Tokyo 100 (JP)
- inventor: Takahashi, Kazuhiro 303 Corpo Omodaka 29 ichijoji-Somedono-cho Sakyo-ku Kyoto City Kyoto Prefecture (JP)

Shibui, Tatsurou M606-11 San White 6-15-4 Minami-Naruse Machida City Tokyo (JP) Kamizono, Michiru F202-24 San White 2-8-11 Tsuruma Machida City Tokyo (JP)

Matsui, Rie A211-23 Sen White 1-26-31 Ogawa Machida City Tokyo (JP)

Teranishi, Yutaka 1-14-2 Takane Sagamihara City Kanagawa Prefecture (JP)

Nakanlahi, Shigetada 517-116 Iwakura-Nagatani-cho Sakyo-ku Kyoto City Kyoto Prefecture (JP)

Kitamura, Naomi 2-272 Yakumo-Higashi-machi Moriguchi City Osaka Prefecture (JP)

Representative: Gutmann, Ernest et al S.C. Ernest Gutmann - Yves Plasseraud 67, boulevard Haussmann F-75008 Paris (FR)

The title of the invention has been amended (Guidelines for Examination in the EPO, A-III, 7.3). The application is published incomplete as filed (Article 93 (2) EPC). The point in the description or the claim(s) at which the omission obviously occurs has been left blank.

- ONA molecules encoding non-A, non-B hepatitis antigens, and their use in producing said antigens.
- Provided herein are a DNA fragment which contains a base sequence coding for a non-A non-B hepatitis-specific antigenic protein occurring in cells of the liver affected with non-A non-B hepatitis, an expression vector in which said DNA fragment is inserted into a cloning site present downstream from a promoter thereof, a transformant obtained by introducing said expression vector into a host, and a process for producing said antigenic protein which comprises providing said expression vector, transforming a host with said expression vector, culturing the transformed host and collecing the protein produced therein.

EP 0 293 274 A1

Bundesdruckerei Berlin

#### Description

5

30

# DNA FRAGMENTS CODING FOR ANTIGENS SPECIFIC TO NON-A NON-B HEPATITIS, EXPRESSION VECTORS CONTAINING SAID DNA FRAGMENTS, TRANSFORMANTS AND PROCESS FOR PRODUCING SAID ANTIGENS

## BACKGROUND OF THE INVENTION

#### Fleid of the invention:

The present invention generally relates to the production of an antigen specific to non-A non-B hepatitis by recombinant DNA technology. More particularly, it relates to a DNA fragment coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis, an expression vector containing such a DNA fragment, a host transformed with such an expression vector, as well as a process for producing said antigen specific to non-A non-B hepatitis by culturing such a transformant.

#### 5 Description of the Prior Art:

Among viral hepatitises, the viral entitles of hepatitis type A and type B have been found and, accordingly, it has now become possible to diagnose such diseases by immunological methods.

Still another type of hepatitis different from the types A and B, which is called non-A non-B type hepatitis, is said to be over 90% of post-transfusion hepatitis: refer to NIPPON RINSHO (Japan Clinic), 35, 2724 (1977); J. Biol. Med., 49, 243 (1976). The pathogenic virus of the non-A non-B type hepatitis, however, has not yet been identified. Only one fact which has already been established is potential infection of human hepatitis type non-A non-B virus to chimpanzee: refer to Lancet I, 459 (1978); ibid., 463 (1978).

Many workers have done various investigations for searching an antigen-antibody system related to the non-A non-B hepatitis by using mainly sera from patients affected with the disease; nevertheless, no definite system has been found. Under these circumstances, the diagnosis of non-A non-B hepatitis should inevitably be effected by so-called exclusion diagnosis: that is, whether or not the hepatitis of a patient is type A or type B or other hepatitis due to a virus known to cause hepatopathy, for example, CMV, HSV, EBV, etc., is first determined; and if not, the patient's hepatitis is diagnosed as non-A non-B type. Thus, such a diagnosis of non-A non-B hepatitis will require much time and labor.

An antigenic protein specific to non-A non-B hepatitis and useful for the direct diagnoses of the hepatitis has been purified from human and chimpanzee hepatocytes affected with non-A non-B hepatitis, and a monoclonal antibody specific to the antigen and useful for the treatment of the non-A non-B hepatitis has also been proposed: refer to Japanese Patent Application Laying-open (KOKAI) Nos. 176856/86 and 56196/86.

A large amount of such an antigenic protein specific to non-A non-B hepatitis should be required when such a protein is to be employed, for example, as a diagnostic agent. However, it is not always appropriate to purify such a large amount of the antigenic protein from chimpanzee hepatocytes affected with non-A non-B hepatitis.

On the other hand, in order to detect a gene coding for a specific antigen of non-A non-B hepatitis by nucleic acid hybridization and, further, to produce such an antigen specific to non-A non-B hepatitis by the recombinant DNA technology, it is essential to obtain a gene fragment coding for the antigenic protein specific to the non-A non-B hepatitis.

## SUMMARY OF THE INVENTION

The present inventors have made great efforts to produce such a specific antigenic protein in a large amount by genetic engineering techniques, and finally isolated a gene fragment coding for the antigenic protein specific to non-A non-B hepatitis, said gene fragment being useful for the production of such antigens. Further, the inventors have successfully constructed an expression vector containing said gene fragment. Thus, the present invention has now been attained.

It is an object of the invention to provide a DNA fragment which contains a base sequence coding for an antigen specifically occurring in a host cell affected with non-A non-B hepatitis or an antigenic protein specific to non-A non-B hepatitis having physiological activities equivalent to those of said specifically occurring antigen.

Another object of the invention is to provide an expression vector having said DNA fragment introduced thereinto at a cloning site downstream from a promoter of the vector.

A still another object of the invention is to provide a transformant obtained by transforming a host cell with said expression vector.

A further object of the invention is to provide a process for producing such an antigen specific to non-A non-B hepatitis by culturing said transformant.

## BRIEF DESCRIPTION OF THE DRAWINGS

Other objects and advantages of the present invention will be apparent from the following detailed description with reference to the attached drawings, in which:

Figs. 1a-1e show the base sequence coding for an antigenic protein specific to non-A non-B hepatitis;

Fig. 2 shows the base sequence of a hybrid promoter Pac;

Figs. 3a-3c show the base sequence of a cDNA fragment obtained in Example 1 described hereinbelow, together with deduced amino acid sequence;

Figs. 4a-4c show the base sequence of cDNA containing the full length gene sequence of an antigenic protein specific to non-A non-B hepatitis, which cDNA was obtained in Example 2 described hereinbelow, the base sequence 57-1388 thereof coding for the antigenic protein specific to non-A non-B hepatitis;

10

15

35

40

50

60

65

Fig. 5 schematically illustrates the construction of a plasmid pCV44H;

Fig. 6 schematically illustrates the construction of a plasmid pCV44B; and

Fig. 7 schematically illustrates the construction of a plasmid pCZ44.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention will be described in detail hereinbelow.

According to one aspect of the invention, a DNA fragment is provided which contains a base sequence coding for an antigenic protein occurring specifically in hepatocytes affected with non-A non-B hepatitis.

Such a DNA fragment of the invention may be prepared in the following manner.

First, a liver tissue specimen derived from a human or chimpanzee individual affected with non-A non-B hepatitis is homogenized in an aqueous solution of guanidinium thiocyanate and then subjected to cesium chloride equilibrium dusity gradient centrifugation according to Chirgwin et al. method (Blochemistry, 18, 5294-5299 (1979)) to separate total RNA as a precipitate. After separation, the total RNA is purified by phenoi extraction and ethanol precipitation.

"Individuals affected with non-A non-B hepatitis" used as sources of liver tissue specimens in the invention may include those affected with so-called type D hepatitis, which has recently been named.

It is known that mRNA of an antigen gene generally has a poly-A chain. Thus, the total RNA is subjected to oligo(dT) cellulose column chromatography in a conventional manner and poly(A)-containing RNA (poly A\*RNA) is isolated as mRNA material.

A cDNA library corresponding to the poly A\* RNA is then obtained from the mRNA material according to the random primer method (Y. Ebina et al., Cell, 40, 747-758 (1980)): Thus, a number of DNAs complementary to the mRNA material are randomly synthesized using any primer of e.g. about 6 bases and a reverse transcriptase.

The cDNA is methylated with a DNA methylase, e.g. <u>EcoRI</u> methylase, to protect cleavage sites present in the cDNA capable of being cleaved by a corresponding restriction enzyme, e.g. <u>EcoRI</u>. A DNA linker containing the corresponding restriction enzyme cleavage sites at both ends, e.g. <u>EcoRI</u> linker (CGAATTCG), is added to the methylated cDNA and, then, this cDNA is digested with the restriction enzyme, e.g. <u>EcoRI</u>.

The digested cDNA is then cloned into a cloning vector such as a plasmid or a  $\lambda$  phage. For example, the cDNA may be introduced into EcoRI site of  $\lambda$ gt 11 DNA, which is an expression cloning vector: refer to R.A. Young et al., Pro. Natl. Acad. Sci. U.S.A., 80, 1194-1198 (1983). The cDNA will be inserted into the  $\beta$ -gal gene on the  $\lambda$ gt 11 phage. Thus, expression of the cDNA can be easily verified by the production of a fused protein with  $\beta$ -galactosidase due to induction of the expression by the lactose operon promoter of said phage when E. coil transfected with said phage is cultured in a medium containing IPTG (isopropytthlo- $\beta$ -D-galactopyranoside).

The  $\lambda gt$  11 phage incorporating the cDNA is then introduced into E. coll by Tomizawa et al. method in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). The thus transfected microorganism is cultured in an IPTG-containing medium.

The thus formed plaques can be easily selected by an immunological screening method using a monoclonal antibody specifically directed to non-A non-B hepatitis to obtain a desired cDNA. Such a monoclonal antibody which can be used in the immunological screening method may be prepared according to the methods described in Japanese Patent Application Laying-open Nos. 176856/86 and 56196/86. The screening methods used may include the western blotting technique described in these applications.

The plaques positive in the immunological screening test are selected to proliferate the phage by Tomizawa et al. method. DNA is purified from the grown phage by T. Maniatis et al. method in "Molecular Cloning", Cold Spring Harbor Laboratory, pp. 85 et seq. (1982), and cleaved with a suitable restriction enzyme such as EcoRI. The thus purified and digested DNA fragments can be used to determine the base sequence of a desired cDNA segment according to Maxam and Gilbert method in Methods in Enzymology, 65, 499-560 (1980); or alternatively, after further cloning the DNA fragments into M13 phage, the base sequence of such a desired cDNA segment can be determined according to the dideoxy method: Sanger et al., Proc. Natl. Acad. Sci. U.S.A., 74, 5463 (1977).

Thus, a cDNA fragment coding for an antigen specific to non-A non-B hepatitis can be obtained. However, such a DNA fragment may usually be only a portion of the gene coding for the non-A non-B hepatitis-specific antigen.

A full length cDNA coding for such a non-A non-B hepatitis-specific antigen may be obtained in the following manner.

Poly A\*-mRNA is Isolated and purified in a manner similar to that described above. From the poly A\*-mRNA a cDNA library is obtained according to Okayama-Berg vector-primer method: Molecular and Cellular Biology, 2, 161-170 (1982).

A plasmid containing such a cDNA thus prepared is used to transform <u>E. coll</u> by any conventional method, for instance, the method D. Hanahan: J. Mol. Biol., <u>166</u>, 557 (1983). The transformant ampicillin-resistant

3

strains are collected and screened by the colony hybridization method using the aforementioned DNA fragment as a probe. Such a probe may preferably be prepared by either the strepto-avidin method, or the nick translation method using photobiotinnucleic acids and <sup>32</sup>P-nucleic acids.

The thus selected colonies containing a cDNA clone are cultured. Plasmid DNA is obtained from the cultured colony according to Birmboim et al. method (Nucleic Acid Res., 7, 1513 (1979)) and digested with a suitable restriction enzyme. The base sequence of a desired full-length cDNA segment is then determined according to the aforementioned Maxam and Gilbert method or, alternatively, after further cloning the digested DNA into M13 phage or pVC12 plasmid, such a base sequence is determined according to the above described Sanger et al. dideoxy method.

The base sequence of the full length DNA coding for an antigen specific to non-A non-B hepatitis is shown in Fig. 1, in which the symbol "-" just under the base sequence represents a corresponding base complementary to the respective base described just above each of the symbols.

Of course, DNA fragments which can be employed in the invention do not necessarily contain the same base sequence as shown in Fig. 1, but those DNA fragments in which a part of said base sequence shown in Fig. 1 has been substituted by at least one different base or deleted therefrom and those DNA fragments in which one or more additional bases have been added to the base sequence of Fig. 1 may also be included herein provided that such different DNA fragments may code for substances having physiological activities equivalent to those of the non-A non-B hepatitis-specific antigens encoded by the base sequence of Fig. 1.

According to another aspect of the invention, an expression vector is provided in which the aforementioned DNA fragment of the invention is inserted into a cloning site downstream from a promoter of this vector.

The expression vector of the invention contains a promoter in a position capable of controlling the transcription of a DNA fragment coding for a non-A non-B hepatitis-specific antigen obtained by the aforementioned method. The promoters used in the invention may be any promoter capable of expressing the DNA fragment in a host, and preferably of controlling the transcription of the fragment.

When a host used is a microorganism such as Escherichia coii, Bacillus subtilis, etc., the expression vector of the invention may preferably comprise a promoter, a ribosome binding sequence, a gene for a non-A non-B hepatitis-specific antigen, a transcription termination factor, and a gene controlling the promoter.

The promoter used may include those derived from <u>E. coli</u>, phage, etc., for example, tryptophan synthase operon (trp), lactose operon (lac), lipoprotein (lpp), recA, lambda phage PL, PR, T5 early gene P25, P26 promoter, which may also be prepared by chemical synthesis. Also included herein are hybrid promoters such as tac (trp:lac), trc (trp:lac) and Pac (phage:E. coli) shown in Fig. 2.

The ribosome binding sequence may be derived from <u>E. coll</u>, phage, etc., but preferably may be those synthetically prepared, for example, those containing a consensus sequence such as

AGGAGGTTTAA.

SD sequence

35

40

The gene for a non-A non-B hepatitis-specific antigen may be directly employed without any modification. Preferably, an unnecessary base sequence (non-coding region) may be deleted by site-directed mutagenesis: BIO TECHNOLOGY, July, 636-639 (1984).

A transcription termination factor may not always be required in the expression vector of the invention. Preferably, the instant vector may contain a p-independent terminator, for example, ipp terminator, trp operon terminator, ribosomal RNA gene terminator, etc.

The expression vector may be derived from any conventional plasmid. Preferably, it may be derived from such a plasmid as replicating itself in <u>E. coli</u> or <u>Bacilius</u> <u>subtilis</u>, for example, pBR322- or pUB110-derived plasmid.

Desirably, these factors required for expression are arranged in the expression plasmid in the order of the promoter, the SD sequence, the structural gene of a non-A non-B hepatitis-specific antigen, and the transcription termination factor from 5' to 3'. A repressor gene required to control the transcription, a marker gene such as drug-resistant gene, and a plasmid replication origin may be arranged in any order in the expression vector.

The expression vector of the invention may be introduced into a host by any conventional method for transformation of E. coli, e.g., one described in Molecular Cloning, 250-253 (1982), or of Bacillus subtilis, e.g., one described in Molec. Gen. Genet., 168, 111-115 (1979) or Proc. Nat. Acad. Sci. U.S.A., 44, 1072-1078 (1958).

The resulting transformant may be cultured in any conventional medium, e.g. one described in Molecular Cloning, 68-73, (1972), at a temperature in the range of 28 to 42°C in both cases of E. coil and Bacilius subtilis. Preferably, it may be cultured at a temperature in the range of 28 to 30°C where no expression of heat shock proteins may be induced.

The desired protein thus produced may be easily purified from the host in conventional procedures. For example, the host cell may be crushed by lysozyme-surfactant or ultra-sonication, and the insoluble fractions which contain the desired non-A non-B hepatitis-specific antigen may be then collected by centrifugation,

4

solubilized in a surfactant such as 0.01% SDS, and subjected to column chromatography using a monoclonal antibody (Japanese Patent Application Laying-open (KOKAI) Nos. 56196/86 and 176856/86.

When an eukaryotic cell such as an animal cell is employed as a host, the expression vector of the invention is preferably as follows:

5

10

20

25

30

40

45

50

55

60

65

The promoters used in the vector of the invention for the expression in eukaryotic cells may herein include SV40 early and late promoters; promoters of apolipoprotein E and A-I genes; promoter of heat shock protein gene (Proc. Natl. Acad. Sci. U.S.A., 78, 7038-7042 (1981)); promoter of metallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6511-6515 (1980)); HSV TK promoter; adenovirus promoter, such as Ad2 major late promoter (Ad2 MLP); LTR (long terminal repeat) of retrovirus; etc. SV40 promoter and promoter of metallothionein gene are preferred.

The expression vector of the invention may contain a splice sequence comprising 5' splice junction donor site, an intron and 3' splice junction acceptor site. A common base sequence is found at all the splice junction sites (exonintron junction sites); so-called GT/AG rule that any intron region always starts from two bases GT at the donor site and terminates at two bases AG of the acceptor site has been established.

The expression vector of the invention may contain one or more splice sequences as mentioned just above. Such splice sequences may be positioned upstream or downstream of the structural gene for a non-A non-B hepatitis-specific antigen.

lilustrative examples of such splice sequences may include those DNA sequences found in exons 2 and 3 of rabbit β-globin gene (Science, 26, 339 (1979)) and mouse methallothionein-I gene containing the promoter, exons 1, 2 and 3 and introns A and B of methallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6513 (1980)). The 5' and 3' splice sites may be derived from the same or different gene; for example, a sequence in which 5' splice site contained in adenovirus DNA is linked to 3' splice site derived from the gene of Ig variable region can be employed.

The expression vector of the invention also contains a polyadenylation site downstream from the structural gene of a non-A non-B hepatitis-specific antigen. Illustrative examples of the polyadenylation sites may include those derived from SV40 DNA,  $\beta$ -globin gene or methallothioneln gene. A combined site of the polyadenylation sites of  $\beta$ -globin gene and SV40 DNA may be employed in the invention.

The expression vector of the invention may also contain a dominant selective marker permitting the selection of transformants. Selective markers which can be used herein may include DHFR gene imparting MTX (methotrexate) resistance to a host; tk gene of herpes simplex virus (HSV) which permits selection of tk strains transformed therewith in HAT medium; the gene for aminoglycoside 3'-phosphotransferase from E. coll transposon Tn5, which imparts to a host the resistance against 3'-deoxystreptamine antibiotic G418; bovine papilloma virus gene permitting morphological discrimination by piled up growth; and aprt gene.

Alternatively, animal cells transformed with the expression vector of the invention may be selected by the cotransformation even though no selective marker is present in the vector. For this purpose, an animal cell is cotransformed with both the expression vector and a plasmid or other DNA containing a gene for such a selective marker and selected by a phenotypic trait of the gene.

Advantageously, the expression vectors may also contain a plasmid fragment having an origin of replication derived from a bacterium such as <u>E. coll</u>, since such vectors can be cloned in bacteria. Such plasmids may include pBR322, pBR327, pML, etc.

Illustrative examples of plasmid vectors used as sources of the expression vectors according to the invention may include pKCR (Proc. Natl. Acad. Sci. U.S.A., 78, 1528 (1981)), which contains SV40 early promoter, the splice sequence and polyadenylation site derived from rabbit  $\beta$ -globin gene, the polyadenylation site from SV40 early region, and the origin of replication and ampicillin resistant gene from pBR322; pKCR H2 (Nature, 307, 605 (1984)), in which the pBR322 portion of pKCR has been substituted by pBR327 fragment and the EcoRI site present in the exon 3 of rabbit  $\beta$ -globin gene has been converted into Hindlil site; and pBPV MT1 containing BPV gene and methallothioneln gene (Proc. Natl. Acad. Sci. U.S.A., 80, 398 (1983)).

Animal cells transformed with the expression vector of the invention may include CHO cells, COS cells, and mouse L cells, C127 cells and FM3A cells.

The introduction of the expression vector of the invention into an animal cell may be carried out by transfection, microinjection, etc. Most often, the transfection may employ CaPO<sub>4</sub>: Virology, <u>52</u>, 456-467 (1973). Animal cells transformed by introducing the expression vector of the invention may be cultured in a

suspension or solid medium by conventional methods. The culture medium used is most often MEM, RPMI1640, etc.

Proteins produced in the transformed animal cells can be separated and purified in the almost same manner as in the case of microorganisms aforementioned.

As stated, the invention provides a transformant cell obtained by introducing the expression vector of the invention into a host cell.

Also provided according to the invention is a process for producing a non-A non-B hepatitis-specific antigen comprising culturing said transformant and collecting the produced and accumulated antigen.

As stated previously, a large amount of an antigenic protein specific to non-A non-B hepatitis is required when such a protein is to be utilized as a direct diagnostic agent. According to the present invention, such an antigenic protein can be produced with a low cost and a large scale without use of infected chimpanzee hepatocytes. Prior to the present invention, it has been difficult obtain such a large amount of a non-A non-B hepatitis-specific antigenic protein from hepatocytes of chimpanzees affected with non-A non-B hepatitis.

Further, the DNA fragment coding for an antigenic protein of non-A non-B hepatitis virus according to the present invention will be useful as a probe for detecting the gene of said antigenic protein by nucleic acid hybridization.

## **EXAMPLES**

10

20

45

65

The following examples will be given by way of illustration but these examples in no way limit the scope of the invention without departing the concept thereof.

EXAMPLE 1: Preparation of cDNA Fragment Coding for Antigenic Protein Specific to Non-A Non-B Hepatitis

Poly(A)-containing RNA was prepared from chimpanzee liver according to the guanidine thiocyanate-lithium chloride method: Cathala et al., DNA, 2, 329 (1983).

The infected liver (5 g) was taken out from a chimpanzee affected with non-A non-B hepatitis and immediately frozen by liquid nitrogen. The frozen liver was added into a Waring blender together with liquid nitrogen and ground at 3,000 rpm for 2 minutes. The ground liver specimen was further ground by a Teflon homogenizer at 5 rpm in 100 ml of a solution: 5 M guanidine thiocyanate, 10 mM EDTA, 50 mM Tris-HCl (pH 7), 80% (v/v) β-mercaptoethanol. The thus solubilized material (20 ml) was slowly placed on 5.7 M CsCl solution (10 ml) contained in a centrifuge tube and centrifuged at 27,000 rpm for 20 hours in Hitachi RPS 28-2 rotor. The thus precipitated RNA was collected and dissolved in 10 ml of a solution: 0.1% sodium laury/sulfate, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5). The RNA was extracted with phenol-chloroform and re covered by ethanol precipitation.

The thus obtained RNA (about 3.95 mg) was dissolved in 1 ml of a solution: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. The solution was incubated at 65°C for 5 minutes, and 5 M NaCl (0.1 ml) was added. The resulting mixture was subjected to chromatography on an oligo(dT) cellulose column (column volume of 0.5 ml, P-L Biochemical). The thus adsorbed poly(A)-containing mRNA was eluted with a solution: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA. There was obtained about 100 µg of poly(A)-containing mRNA.

The thus obtained poly(A)\* mRNA (10 µg) was dissolved in 50 µ of RT buffer: 20 mM Tris-HCl (pH 8.8), 0.1 M KCl, 12 mM MgCl<sub>2</sub>, 2 mM MnCl<sub>2</sub>. To this solution, there was added 8 µg of random primer d(N)s (P-L Biochemical). The resulting mixture was heated at 95°C for 3 minutes to denature the materials, which was then cooled gradually to room temperature to anneal the random primer with the mRNA. To the annealed mixture, there were aded 10 mM 4NTP (10 µl) and reverse transcriptase (225 units) from TAKARA SHUZO (Japan), and then water was added so as to make the total volume of the mixture to 100 µl. Reaction was allowed to proceed at 42°C for one hour.

To the reaction mixture (50  $\mu$ ), there were added 10 mM NAD (2  $\mu$ ), 10 mM 4dNTP (10  $\mu$ ), RNase H (5 units), E. coll ligase (1 unit), E. coll DNA polymerase I (6.3 units), and 10  $\times$  T4 DNA ligase buffer (10  $\mu$ ); 0.1 M Tris-HCl, pH 7.5, 0.1 M DTT, 60 mM MgCl<sub>2</sub>) to make the total volume to 100  $\mu$ l. The mixture was allowed to react at 37°C for one hour to synthesize a double stranded DNA.

The thus obtained double stranded DNA was extracted with an equal volume of water-saturated phenol. Phenol in the aqueous layer was removed with the aid of ether followed by ethanol precipitation. The precipitate thus obtained was dissolved in 50  $\mu$ l of water, and 10  $\times$  T4 DNA polymerase buffer (10  $\mu$ l; 0.33 M Tris-acetic acid, pH 7.9, 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM DTT), 10 mM 4dNTP (10  $\mu$ l), and T4 DNA polymerase (6 units) were added to make the total volume to 100  $\mu$ l. The mixture was reacted at 37° C for one hour. There was obtained a double stranded DNA having blunt ends, which was then extracted with phenol to remove proteins and purified by ethanol precipitation as described above. The thus purified DNA was then air dried.

To the purified DNA, there were added 50 mM Tris-HCl (pH 7.5), 1 mM Na<sub>2</sub>EDTA, 5 mM DTT (20  $\mu$ ), 100  $\mu$ M S-adenosyl-L-methionine (2  $\mu$ l), and 1.8 mg/ml EcoRl methylase (0.2  $\mu$ l). Reaction was effected at 37°C for 15 minutes, whereby methylating the EcoRl restriction enzyme cleavage site on the DNA fragment. The reaction mixture was then heated at 70°C for 15 minutes to deactivate the enzyme.

To the reaction mixture, there was added 3'-phosphorylated EcoRl linker (GGAATTCC) in an amount of 100 molecules thereof per molecule of the synthetic DNA. There were further added  $10 \times T4$  DNA ligase buffer (5  $\mu$ ); 0.5 M Tris-HCl, pH 7.5, 60 mM MgCl<sub>2</sub>, 10 mM DTT), 0.1 M ATP (5  $\mu$ ), and T4 DNA ligase (5 units) to make the total volume to 50  $\mu$ l. The resulting reaction mbxture was reacted at 4°C for 16 hours followed by heating at 70°C for 10 minutes to deactivate the enzyme. Then,  $10 \times EcoRl$  buffer (10  $\mu$ ); 15 M Tris-HCl, pH 7.5, 0.5 M NaCl, 60 mM MgCl<sub>2</sub>), and EcoRl (100 units) were added to make the total volume to 100  $\mu$ l, and the reaction mixture was reacted at 37°C for 2 hours to cut the linker. The reaction mixture was passed through Bio Gel A-50 (0.2 cm  $\times$  32 cm, Bio RAD). Elution was effected by a buffer: 10 mM Tris-HCl (pH, 7.5), 6 mM MgCl<sub>2</sub>. Excess EcoRl linker was removed and, thus, a double stranded cDNA having EcoRl sites at both ends thereof was purified.

To the thus obtained double stranded cDNA fragment having EcoRI sites at both ends, there were added gt 11 DNA (10  $\mu$ g) cleaved with EcoRI, 10 × T4 DNA ligase buffer (10  $\mu$ l) as described above, 0.1 M ATP (10  $\mu$ l), and T4 DNA ligase (10 units) to make the total volume to 100  $\mu$ l. The mixture was reacted at 4°C for 16 hours. Thus, said double stranded cDNA fragment was inserted into  $\lambda$ gt 11 DNA.

The  $\lambda$  phage packaging kit (PROMEGA, Biotech) was used to introduce said DNA into  $\lambda$  phage particle. The procedures for packaging were effected according to the instructions of the kit.

The Agt 11 phage having said DNA packaged thereinto was used to transfect E. coli strain Y1090 to form

plaques according to the conventional Tomizawa et al. methods described in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by twanami Shoten (Japan). Among about 200,000 plaques, one positive clone was selected by Immunological screening as described hereinbelow. A monoclonal antibody used in the immunological screening was prepared by the method described in Japanese Patent Application Laying-open (KOKAI) No. 176856/86.

E. coll Y1090 (R.A. Young et al., Pro. Natl. Acad. Sci. U.S.A., 80, 1194-1198 (1983), which had been transfected with \(\lambda\)gt 11, was inoculated in a petri dish together with soft agar held at 42°C. The transfected cell was allowed to stand at 42° C for 5 hours. A nitrocellulose filter (S & S, BA-83, pore size of 0.2 μm) containing 10 mM IPTG was placed on the cell in the dish and incubation was effected at 37°C for 3-4 hours. This nitrocellulose filter was lightly rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed in the TBS buffer (400 ml) containing 3% gelatine and shaked at 40°C for one hour. Thus, the nitrocellulose filter was blocked. Then, a monoclonal antibody (OD280 = 4.3) directed to a non-A non-B hepatitis-specific antigen was added to TBS buffer containing 1% gelatine with a dilution of 1/400. This mixture was put into a vinyi bag together with the filter in a proportion of 2 ml of the mixture per filter, and reaction was allowed to proceed at room temperature for 16 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes. A labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad) was added to TBS buffer containing 1% gelatine with a dilution of 1/1,000. This mixture and the filter were put into a vinyl bag with a proportion of 2 ml of the mixture per filter. Reaction was allowed to proceed at room temperature for 2 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes, in the same manner as described above. Color development was effected by dipping the filter and 4-chloro-1-naphthol (12 mg, Bio Rad) into 20 ml of TBS buffer containing hydrogen peroxide. After completion of the color development, the filter was thoroughly washed with water and put into a vinyl bag containing water. The bag was stored in a dark and cold place.

10

20

30

45

50

55

65

Thus, one positive plaque was obtained. The plaque was three times subjected to single plaque isolation. In each time, immunological screening was effected in the same manner as described above, verifying that the plaque was in fact positive.

The phage was then cultured in a large scale to purify the DNA in the following manner: First, E. coll Y1090 was cultured overnight in 10 ml of NZ medium prepared by adding NZ amine (10 g), NaCl (5 g) and 5 mM MgCl<sub>2</sub> to one liter of water followed by adjusting the pH to 7.2. The culture (1 ml) was transfected with the phage, with the m.o.l. (multiplicity of infection) being 0.1. The transfected culture was allowed to stand at 37°C for 10 minutes and then transferred to one liter of NZ medium. Shaking culture was effected at 37°C for 7-8 hours until the cells were lysed. Chloroform (5 ml) was added to the culture and shaking was continued for additional 30 minutes. The culture was subjected to centrifugation at 6,500 rpm for 10 minutes to remove cell debris.

NaCl (29 g) and polyethylene glycol (70 g) were added to and thoroughly dissolved in the obtained supernatant, and the solution was allowed to stand at 4°C overnight. The precipitate was collected by centrifugation at 8,500 rpm for 20 minutes, drained thoroughly, and dissolved in 20 ml of TM buffer: 10 mM Tris-HCl (pH 7.5), 5 mM MgCl<sub>2</sub> DNase I and RNase A were added to the solution, both with a concentration of 10 μg/ml, and the reaction was effected at 37°C for one hour. Chloroform (20 ml) was then added to the reaction mixture and stirred; thus, polyethylene glycol was distributed in the chloroform layer which was then separated from the aqueous layer. This aqueous layer was ultra- centrifuged at 28,000 rpm for 60 minutes. Thus, a pellet of phage particles was obtained.

This pellet was dissolved in TM buffer (1 ml) and subjected to CsCl density gradient centrifugation at 33.000 rpm for 20 hours. The resultant fraction containing the phage particles ( $\rho=1.45$ -1.50) was dialyzed overnight against TM buffer. Proteinase K was added to the dialyzate in an amount of 100  $\mu$ g/ml and reaction was effected at 37°C for one hour. Thereafter, an equal volume of water-saturated phenol was added and phenol-extraction was gently effected. After centrifugation at 6,500 rpm for 10 minutes, the aqueous layer was removed, put into a dialysis tube, and dialyzed overnight against water at 4°C. Thus, about 5 mg of DNA was obtained.

Cleavage reaction of this DNA (100 μg) with EcoRi (100 units) in the aforementioned buffer (100 μl) at 37°C revealed that two cDNA segments of 390 bp and 345 bp were inserted into the phage DNA.

These two EcoRI fragments were re-cloned into EcoRI site of a cloning vector pUC 119. Base sequences of these DNA fragments were determined by the dideoxy method using commercially available primers CAGGAAACAGCTATGAC and AGTCACGACGTTGTA, respectively. The base sequence of the linking portion between these two DNA fragments was similarly determined by cutting this cDNA fragment at BamHI and EcoRV sites present therein with corresponding specific restriction enzymes, inserting the resulting BamHI-EcoRV DNA fragment between BamHI and Small sites of the plasmid pUC 119, and sequencing the fragment by the dideoxy method.

The base sequence of said cDNA fragment is shown in Fig. 3. This was a partial cDNA fragment of a gene coding for an antigenic protein specific to non-A non-B hepatitis.

# EXAMPLE 2: Preparation of cDNA Containing the Full Length Gene Sequence

Messenger RNA was prepared as described in Example 1 and cDNA was synthesized using Okayama vector according to the conventional method described in Molecular Cloning, p. 211 et seq. The procedures used to synthesize cDNA were as follows:

To 300 µl of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 10 mM NaCl), there were added 400 µg of

pCDV 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 500 units of Kpni (TAKARA SHUZO, Japan), all restriction enzymes used hereinafter having been manufactured by TAKARA SHUZO (Japan) unless otherwise noted. Reaction was effected at 37°C for 6 hours to cut the plasmid at Kpni site therein. After phenol-chloroform extraction, ethanol precipitation was effected to recover DNA.

The DNA (about 200 μg) cleaved with <u>Kpnl</u> was added to 200 μl of a solution which was obtained by adding dTTP in a concentration of 0.25 mM to a buffer (TdT buffer): 40 mM sodium cacodylate, 30 mM Tris-HCl (pH 6.8), 1 mM CaCl<sub>2</sub>, 0.1 mM dithiothreitol (DTT). Further, 81 units of terminal deoxynucleotidyl transferase (TdT. manufactured by P-L Blochemicals) was also added. Reaction was effected at 37°C for 11 minutes. Thus, a poly(dT) chain (about 67 deoxythymidylic acid residues) was added to the 3′ end at the <u>Kpnl-cleaved</u> site of pCDV 1. After phenol-chloroform extraction and ethanol-precipitation, about 100 μg of pCDV 1 DNA to which poly(dT) chain had been added was recovered from the reaction mixture.

The thus obtained DNA was added to 150 μl of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 100 mM NaCl), and <u>Hpal</u> (360 units) was also added, followed by reaction at 37°C for 2 hours. The reaction mixture was subjected to electrophoresis on agarose gel to separate and recover about 3.1 Kbp DNA fragment. Thus, there was obtained about 60 μg of poly(dT)-containing pCDV 1.

The thus obtained DNA was dissolved in 500  $\mu$ l of a solution (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), incubated at 65°C for 5 minutes, and cooled on ice. After adding 5 M NaCl (50  $\mu$ l), the mixture was subjected to chromatography on oligo(dA) cellulose column (Colaborative Research). DNA having a poly(dT) chain of sufficient length was adsorbed on the column and eluted with a solution: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Thus, there was obtained 27  $\mu$ g of pCDV 1 to which poly(dT) chain had been added, abbreviated hereinafter as vector primer.

A linker DNA was prepared in the following manner: To 200  $\mu$ l of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 50 mM NaCl), there were added about 14  $\mu$ g of pL 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 50 units of Pstl. Reaction was effected at 37°C for 4 hours to cut the pL 1 DNA at Pstl site. Phenol-chloroform extraction and ethanol precipitation of the reaction product gave about 13  $\mu$ g of pL 1 DNA cleaved at Pstl site.

The thus obtained DNA (about 13  $\mu$ g) was added to 50  $\mu$ l of the TdT buffer containing dGTP at a final concentration of 0.25 mM, and 54 units of TdT (P-L Biochemicals) was also added. The mixture was incubated at 37°C for 13 minutes to add a (dG) chain (about 14 deoxyguanylic acid residues) to the 3' end at the Pstl-cleaved site of pL 1. After phenol-chloroform extraction, DNA was recovered by ethanol precipitation.

The thus obtained DNA was added to 100 µl of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 60 mM NaCl), and 80 units of <u>Hindlll</u> was also added. The mixture was incubated at 37°C for 3 hours to cut the pL 1 DNA at <u>Hindlll</u> site. The reaction product was fractionated by agarose gel electrophoresis. About 0.5 Kb DNA fragment was recovered by the DEAE paper method: Dretzen et al., Anal. Biochem., 112, 295 (1981). Thus, there was obtained an oligo(dG) chain-containing linker DNA, hereinafter abbreviated simply as linker DNA.

The aforementioned poly(A)+ RNA (about 2 µg) prepared in the same manner as in Example 1 and the vector primer (about 1.4 µg) were dissolved in 22.3 µl of a solution: 50 mM Tris-HCl (pH 8.3), 8 mM MgCl<sub>2</sub>, 30 mM KCl, 0.3 mM DTT, 2 mM dNTP (dATP, dTTP, dGTP and dCTP) and 10 units of ribonuclease inhibitor (P-L Biochemicals). To the solution, there was added 10 units of reverse transcriptase manufactured by SEIKAGAKU KOGYO (Japan). Incubation was effected at 37°C for 40 minutes to synthesize a DNA complementary to the mRNA. After phenol-chloroform extraction and ethanol precipitation, the vector primer DNA to which a double stranded RNA-DNA had been added was recovered.

The thus obtained vector primer DNA containing RNA-DNA double stranded chain was dissolved in 20  $\mu$ i of TdT buffer containing 60  $\mu$ M dCTP and 0.2  $\mu$ g poly(A). After adding 14 units of TdT (P-L Biochemical), the mixture was incubated at 37°C for 8 hours to add a (dC) chain of 12 deoxycytidytic acid residues to the 3′ end of the cDNA. The reaction product was extracted with phenol-chloroform and precipitated with ethanol to recover a cDNA-vector primer DNA to which a (dC) chain had been added.

The thus obtained (dC) chain-containing cDNA-vector primer DNA was dissolved in 400 µl of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 60 mM NaCl), and 20 units of Hindlil was also added. The mixture was incubated at 37°C for 2 hours to cut the DNA at Hindlil site. The reaction product was extracted with phenol-chloroform and precipitated with ethanol. Thus, there was obtained 0.5 pmole of a (dC) chain-containing cDNA-vector primer DNA.

The thus obtained (dC) chain-containing cDNA-vector primer DNA (0.08 pmole) and the aforementioned linker DNA (0.16 pmole) were dissolved in 40  $\mu$ l of a solution: 10 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 1 mM EDTA. The resulting solution was incubated at 65°C for 10 minutes, at 42°C for 25 minutes, and then at 0°C for 30 minutes. The reaction mixture was adjusted to 20 mM Tris-HCl (pH 7.5), 4 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M KCl and 0.1 mM  $\beta$ -NAD in a total volume of 400  $\mu$ l.

To the reaction mixture, there was added 10 units of <u>E. coli</u> DNA ligase (New England Biolabs), followed by incubation overnight at 11°C. After adjusting the concentrations of dNTP and β-NAD in the reaction mixture to 40 μM and 0.15 mM, respectively, by supplementing necessary reagents, 5 units of <u>E. coli</u> DNA ligase, 7 units of <u>E. coli</u> DNA polymerase I (P-L Biochemicals) and 2 units of <u>E. coli</u> ribonuclease H (P-L Biochemicals) were added to the reaction mixture. The mixture was incubated at 12°C for one hour and then at 25°C for one hour.

In the course of the above reactions, a recombinant DNA containing the cDNA was cyclized and the RNA portion of the RNA-DNA double stranded chain was substituted by DNA. Thus, a desired recombinant plasmid containing a complete double-stranded DNA was produced.

The recombinant plasmid was used to transform competent cells of E. coli strain MC1064 prepared by conventional methods. Approximately 50,000 transformants were fixed on a nitrocellulose filter. These colonies were screened according to the colony hybridization method described in Molecular Cloning, Cold Spring Harbor Laboratory, p. 329 et seq. (1982) using the cDNA fragment obtained in Example 1 as a 32P-labelled probe. Thus, three clones showed strong hybridization at 42°C.  These positive clones were analyzed in detail by Southern method: J. Mol. Biol., 98, 503 (1975). There was obtained the desired full length cDNA of a gene coding for an antigenic protein specific to non-A non-B hepatitis. The base sequence of the cDNA is shown in Fig. 4.  The expression vector containing the full length cDNA was designated as pCDVCL-I.  EXAMPLE 3: Preparation of Expression Vector and Transformant and Expression of Specific Antigen	10
A. Preparation of Expression Vector and Transformant	
l) Modification of N-terminus (Fig. 5):  i) In 100 μl of a buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl <sub>2</sub> ), pCDVCL-I (5 μg) was digested with Pvul (10 units) at 37°C for 2 hours. The reaction mixture was heated at 75°C for 15 minutes to deactivate the enzyme, diaprace and dried. The cleaved plasmid DNA was treated with T4 DNA polymerase (4 units) in 40 ul set another. 30 m 14 Trib	15
DNA polymerase (4 units) in 40 µl of a system: 33 mM Tris-acetic acid (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate and 0.5 mM dithiothreitol, to which 2 mM 4-deoxytriphosphate had been added; thus, the 3' protruding end of the plasmid DNA was filled in to produce a blunt end. The thus treated mixture was heated at 70°C for 10 minutes to deactivate the enzyme, dialyzed against water, and dried. The thus obtained plasmid DNA was then stored in the form of an aqueous solution (50 µl). This plasmid DNA fragment is hereinafter designated as Fragment I.	20
ii) On the other hand, pCDVCL-I (20 μg) was digested with Ncol and Hindlil (each 20 units) at 37°C for 2 hours in 100 μl of a buffer: 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 6 mM MgCl <sub>2</sub> . The plasmid DNA was subjected to 5% acrylamide gel electrophoresis at 10 V/cm for 1.5 hours in a buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA. The gel was stained with 0.05% aqueous ethicium bromide solution and two gel slices corresponding to DNA fragments of larger molecular weights were excited from the relationship of the relation	25
of a buffer for DNA extraction (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% sodium laurylsulfate), and allowed to stand overnight at 37°C to extract DNA from the gei. The materials were subjected to centrifugation at 10,000 rpm for 15 minutes to eliminate larger gel pieces, and passed through a glass filter to remove smaller gel pieces. The DNA was purified by effecting attempt possibled.	30
three times and stored in the form of an aqueous solution (200 µl). This plasmid DNA fragment is hereinafter designated as Fragment II.  iii) A primer of the DNA portion to be modified as shown below (51 bases) was synthesized by a DNA synthesizer, NIKKAKI (Japan), Applied Blosystem MODEL 380A. The synthesized DNA was overnight reacted with concentrated aqueous ammonia at 55°C to deprotect and purified by reversed HPLC before use.	<i>35</i>
Primer ACAACAGATCTAAGCTTATGGCAGTTACAACAAGATTAA	45
X X X X X X (Original sequence) (A-A-GGTC-CG-	~
. CATGGTTGCATG wherein x represents a	50
base substitution.	
•	55
The synthetic primer (150 pmole) was treated with T4 polynucleotide kinase (20 units) in 10 µl of a kinase buffer (50 mM Tris-HCl, pH 8.0, 10 mM MgCl <sub>2</sub> , 5 mM dithiothreitol) to phosphorylate the 5' end thereof.  iv) Fragment I (0.05 pmole), Fragment II (0.05 pmole) and 5'-phosphorylated primer (45 pmole) were added to 12 µl of 5 × polymerase-ligase buffer (0.5 M NaCl, 32.5 mM Tris-HCl, pH 7.5, 40 mM MgCl <sub>2</sub> , 5 mM	60
β-mercaptoethanol) to make the total volume of the mixture 34.8 μl. The mixture was boiled at 100°C for 3 minutes, immediately after which it was placed in a thermostat at 30°C and allowed to stand for 30 minutes. The mixture was allowed to stand at 4°C for 30 minutes and then on ice for 10 minutes to form a heteroduplex.	-

To an aqueous solution (11.6 μl) containing the heteroduplex, there were added 2.5 mM 4-deoxynucleotide

triphosphate (2  $\mu$ ), 10 mM ATP (2  $\mu$ ), Klenow enzyme (2 units) and T4 DNA ligase (0.5 units) to form a mixture of 20  $\mu$ l in total volume. The mixture was reacted overnight at 16°C to cyclize the DNA.

An aqueous solution (2 µl) containing the circular DNA was used to transform E. coll HB101 strain according to conventional methods. Plasmids were separated from the transformant and purified in conventional manners. The plasmid was cleaved with restriction enzyme Hindill and subjected to 5% acrylamide gel electrophoresis. Thus, two separate fragments were collected as desired modified, variant plasmids. Since resulting variant plasmids might often be admixed with original wild-type plasmids, the thus obtained variant plasmids were again employed to transform E. coll HB101 so as to purify the plasmid.

Thus, a purified plasmid pCV44H was obtained (Fig. 5).

II) Modification of C-terminus (Fig. 6):

i) Plasmid pCDVCL-I (5  $\mu g$ ) was treated in the same manner as in I) i) described above to produce Fragment I.

ii) Plasmid pCDVCL-I (20 µg) was treated in the same manner as in I) ii) described above except that Ncol and Nsii (each 5 units) were employed. Thus, Fragment II was produced.

iii) In the same manner as in i) iii) described above, the following primer (46 bases) was synthesized and the 5' end thereof was phosphorylated.

20

25

30

35

40

45

50

55

60

10

15

Primer GCACAAGGAAAAAATGAGATCTGTCGACGGTTCACGTA

\*

(Original sequence) (------AGATATGTGAA\*A-----
AATTTCC wherein x represents a base substitution and \* represents ------) an addition.

iv) The Fragment I and II and the 5'-phosphorylated primer obtained above in II) i) to III) were treated in the same manner as in I) iv) described above. Thus, plasmid pCV44B was obtained (Fig. 6).

iii) introduction of cDNA coding for specific antigen into expression vector (Fig. 7):

i) In 100 μl of a buffer H (10 mM Tris-HCi, pH 7.5, 100 mM NaCl, 6 mM MgCl<sub>2</sub>), 10 μg (about 3 pmole) of pCV44H was cut with Hindill (20 units) and Sacl (20 units) at 37° C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis. Thus, a 467 bp DNA fragment coding for the N-terminus of the specific antigen was separated and purified. This fragment is hereinafter designated as Fragment N.

ii) In 100 μl of the buffer H, 10 μg (about 3 pmole) of pCV448 was cleaved with Bgill (20 units) and Saci (20 units) at 37°C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis to isolate and purify a 836 bp DNA fragment coding for the C-terminus of the specific antigen. The thus obtained fragment is hereinafter designated as Fragment C.

iii) In 20 μl of buffer H, 2 μg (about 1 pmole) of an expression vector pUSΔH was cut with Hindlii (2 units) and Bgili (2 units) at 37°C for 2 hours. The reaction mixture was extracted with an equal volume of water-saturated phenol to remove proteins. After extracting the phenol with ether, the reaction mixture was dialyzed against water to desait, and concentrated by a vacuum pump. Thus, there was obtained 10 μl of an aqueous solution containing an expression vector fragment HB.

iv) Fragment N (0.5 pmole), Fragment C (0.5 pmole) and the expression vector fragment HB (0.1 pmole) were mixed and reacted with T4 DNA ligase (1 unit) at 4°C for 16 hours in 10 μl of a buffer (10 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, 6 mM MgCl<sub>2</sub>, 1 mM ATP). The reaction mixture (3 μl) was used to transform commercially available <u>E. coli</u> JM109 competent cell according to conventional methods. The resulting transformants were selected in L broth plate (bactopeptone 10 g, yeast exstract 5 g, NaCl 10 g, agar 15 g per liter) containing 20 μg/ml ampicillin. Thus, there was obtained an expression vector pCZ44 containing the specific antigen gene inserted thereinto (Fig. 7).

B. Expression of Specific Antigen

E. coli strain JM109 possessing pCZ44 was cultured overnight at 30°C in L broth. The culture was inoculated in a fresh L broth with a dilution of 1/50 and cultured with shaking at 30°C for 2 hours. After IPTG (isopropylthio-β-D-galactopyranoside) was added to the medium in a concentration of 2 mM, shaking culture was continued at 30°C for further 3 hours. The cells were collected by centrifugation at 6,500 rpm for 10 minutes and suspended in a buffer (0.9% NaCl, 10 mM Tris-HCl, pH 7.5) to store.

65

C. Verification of Expression of Specific Antigen

The thus obtained cell culture (0.3 ml) was subjected to 10% SDS polyacrylamide gel electrophoresis at 120 V for one hour in a buffer (Tris 3g/1, glycine 14.4 g/1, 0.1% SDS). The gel was removed, placed on a nitrocellulose filter, interposed between filter papers and electrophoresed at 5 V/cm, 4°C in a buffer (Tris 3g/1, glycine 14.4 g/1) to transferred proteins in the gel onto the nitrocellulose filter. The nitrocellulose filter was rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed into 400 ml of TBS buffer containing 3% gelatine and shaked at 40°C for one hour to block the nitrocellulose filter.

To TBS buffer containing 1% gelatine, there was added a monoclonal antibody directed to a non-A non-B hepatitis-specific antigen (OD<sub>280</sub> = 4.3) with a dilution of 1/400. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 16 hours. The reaction mixture was washed three times with 400 ml of TBS buffer containing 0.05% Tween 20 for 10 minutes.

To TBS buffer containing 1% gelatine, there was added a labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad), with a dilution of 1/1000. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 2 hours. The reaction mixture was washed three times with 400 ml of TBS buffer containing 0.05% Tween 20 for 10 minutes.

Color formation was effected by immersing the filter into 20 ml of TBS buffer containing 12 mg of 4-chloro-1-naphthol (Bio Rad) and hydrogen peroxide. After completion of color formation, the filter was thoroughly washed with water, put into a vinyl bag containing water, and stored in a dark and cold place. Such a test effected showed that a protein reacting with

The monoclonal antibody was found at the same position (44 Kd) as found in the case of the specific antigen derived from infected chimpanzee liver. This verifies that such a specific antigen can be in fact expressed in E coil. The invention thus also relates to a process for the in vitro diagnosis of NON-A NON-B hepatitis, which comprises contacting a liver sample and/or a serum sample taken from a patient possibly infected with a NON-A NON-B hepatitis with the protein whose formula appears in claim 3 hereafter or a part thereof for a time and under conditions sufficient to allow for the production of a complex between said protein or part thereof with the antibodies contained in the patient sample and detecting the presence of the immunologic complex, particularly when the patient is suffering from NON-A NON-B hepatitis.

Any part of said protein,or any recombinant, produced by genetic engineering and including the aminoacid sequence of said protein or part of said protein can be substituted for above-said protein, it being understood that the said recombinant protein or part of said protein are specifically recognized by the same antibodies as those which recognize said protein.

In other words the invention relates to all recombinant proteins or protein fragments which bind to antibodies contained in a liver extract or serum sample, or both, and originating from a patient suffering from NON-A NON-B hepatitis.

The invention also relates to a process for detecting in vitro an infection by a NON-A NON-B hepatitis virus, which process comprises contacting the DNA of claim 4, or a fragment thereof, under suitable hybridization conditions, with a sample of liver extract and/or serum sample originating from the patient to be diagnosed and in which the nucleic acid components had previously been made accessible to hybridization, to form a hybridization product between said DNA of claim (probe) and the viral DNA of a NON-A NON-B hepatitis B virus, and detecting said hybridization product, particularly in the case where the patient is indeed infected with a NON-A NON-B virus.

#### Claims

- A DNA fragment which contains a base sequence coding for a non-A non-B hepatitis-specific antigenic protein occurring in cells of the liver affected with non-A non-B hepatitis.
- The DNA fragment in accordance with claim 1, in which the cells of the liver are derived from a human or chimpanzee individual.
- 3. The DNA fragment in accordance with claim 1, in which the antigenic protein specific to non-A non-B hepatitis has the whole or a part of the amino acid sequence represented by the formula:

20

25

30

35

40

45

50

55

60

Met Ala Val Thr Thr Arg Leu Thr Trp Leu His Glu Lys Ile Leu Gln Asn His Phe Gly Gly Lys Arg Leu Ser Leu Leu Tyr Lys Gly Ser Val His Gly Phe His Asn Gly Val Leu Leu Asp Arg Cys Cys Asn Gln Gly Pro Thr Leu Thr Val Ile Tyr Ser Glu Asp His Ile Ile Gly Ala Tyr Ala Glu Glu Gly Tyr Gln Glu Arg Lys Tyr Ala Ser Ile Ile Leu Phe Ala Leu Gln Glu Thr Lys Ile Ser Glu Trp Lys Leu Gly Leu Tyr Thr Pro Glu Thr Leu Phe Cys Cys Asp Val Ala Lys Tyr Asn Ser Pro Thr Asn Phe Gln Ile Asp Gly Arg Asn

130 Arg Lys Val Ile Met Asp Leu Lys Thr Met Glu Asn Leu Gly Leu Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg Cys Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu 170 180 Leu Arg Lys Ser Leu Leu Ser Ala Leu Arg Thr Tyr Glu Pro Tyr Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu Gly Pro Ile 200 210 Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr 230 240 Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu 260 270 Ser Glu Lys Glu Gly Gly Leu Cys Met Asp Asp Ile Ser Tyr Ile Leu Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu 290 300 Ser Ile Lys Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu Lys Asp Arg Ile His Cys Val Ala Phe Val Phe Asp Ala Ser Ser 330 320 Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile Arg Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu 350 360 Thr His Val Asp Ser Met Asp Leu Ile Thr Lys Gly Asp Leu Ile Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys Leu Glu Glu Val 380 390 Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser

Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu

Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala Ala Asp Asp Phe

Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu Arg Glu Glu

Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*\*.

4. The DNA fragment in accordance with claim 1, in which the base sequence comprises the whole or a part of the base sequence represented by the formula:

			1	LO			20			30			4	10
5'	ATG	GCA	GTG	ACA	ACT	CGT	TTG	ACA	TGG	TTG	CAT	GAA	AAG	ATC
3'														
			- ^			60				7.0			90	
	ama.	<b>~~</b>	20	C a m	mmm	60	ccc	220	ccc	COO	AGC.	Cmm	CTC	ጥልጥ
			AAT											INI
		90		_	10	00			110			120		
	AAG	GGT	AGT	GTC	CAT	GGA	TTC	CAT	AAT	GGA	GTT	TTG	CTT	GAC
-														
											•			
	13	30			140			150		•	10	50		
			TGT											AGT
				100			3 (	20			200			210
1	170	61 m	C N TH	180	mm	CCA	19	90 mam	CCX	CAA	200	CCT	ሞልሮ	210 CAG
1	170 GAA	GAT	CAT	ATT	ATT	GGA	19 GCA	90 TAT	GCA	GAA	200 GAG	GGT	TAC	210 CAG
1	170 GAA	GAT	CAT	ATT	ATT	GGA	GCA	90 TAT	GCA	GAA	200 GAG	GGT	TAC	210 CAG
1	L70 GAA	GAT	CAT	ATT	ATT	GGA	GCA	TAT	GCA	GAA	GAG	GGT	TAC	CAG
1	GAA	GAT	CAT	ATT 	ATT	GGA	GCA  230	TAT	GCA	GAA  240	GAG	GGT	TAC	CAG 
1	GAA	GAT 	CAT	ATT 20 TAT	ATT  GCT	GGA 	GCA  230 ATC	TAT	GCA  CTT	GAA  240 TTT	GAG  GCA	GGT 	TAC	CAG 
1	GAA	GAT 	2: AAG	ATT 20 TAT	ATT GCT	GGA TCC	GCA  230 ATC	ATC	GCA  CTT	GAA 240 TTT	GAG  GCA	GGT  CTT	2: CAA	CAG  50 GAG
1	GAA GAA	AGA	2: AAG	ATT  20 TAT	GCT	TCC	GCA  230 ATC 	ATC	GCA  CTT 	GAA  240 TTT 	GAG GCA	CTT	TAC  2: CAA 	CAG  50 GAG 
1	GAA GAA	AGA	2: AAG	ATT  20 TAT	GCT	TCC	GCA  230 ATC 	ATC	GCA  CTT 	GAA  240 TTT 	GAG GCA	CTT	TAC  2: CAA 	CAG  50 GAG 
1	GAA GAA	AGA	2: AAG	ATT  20 TAT	GCT	TCC	GCA  230 ATC 	ATC	GCA  CTT  2 GGA	GAA 240 TTT B0 CTA	GAG GCA 	CTT	TAC  CAA  290 CCA	CAG  50 GAG 
1	GAA GAA	AGA	2: AAG  260 ATT	ATT 20 TAT TCA	GCT 	TCC 270 TGG	GCA  230 ATC 	ATC	CTT 2 GGA	GAA  240 TTT  B0 CTA	GCA 	CTT	290 CCAA	CAG  50 GAG 
	GAA  GAA 	AGA AAA	2: AAG  260 ATT	ATT 20 TAT TCA	GCT	GGA TCC  270 TGG 	GCA  230 ATC 	ATC	GCA CTT 2 GGA	GAA  240 TTT  80 CTA	GCA 	CTT ACA	290 CCAA	CAG  50 GAG  GAA
	GAA GAA ACT ACA	AGA AAA AAA CTG	2: AAG  260 ATT	ATT 20 TAT TCA	GCT GAA	GGA TCC TCC TGG TGG TGG TGG	GCA 230 ATC AAA	ATC CTA	GCA CTT 2 GGA 320 AAA	GAA 240 TTT 80 CTA	GAG GCA TAT	CTT ACA 330 TCC	290 CCA	CAG  50 GAG  GAA

34	10		3	350			360			37	70		
AAT	TTC	CAG	ATA	GAT	GGA	AGA	AAT	AGA	AAA	GTG	ATT	ATG	GAC
380			390			40	0		4	110			420
TTA	AAG	ACA	ATG	GAA	AAT	CTT	GGA	CTT	GCT	CAA	AAT	TGT	ACT
		43	30		4	40			450			46	50
ATC	TCT	ATT	CAG	GAT	TAT	GAA	GTT	TTT			GAA	GAT	TCA
								'-					
	4	170			480	-		49	90		5	500	
CTG	GAC	GAA	AGA	AAG	ATA	AAA	GGG	GTC	ATT	GAG	CTC	AGG	AAG
	510			52	20		5	30			540		
AGC	TTA	CTG	TCT	GCC	TTG	AGA	ACT	TAT	GAA	CCA	TAT	GGA	TCC
5	50 ·			560			570			58	30		
CTG	GTT	CAA	CAA	ATA	CGA	ATT	CTG	CTG	CTG	GGT	CCA	ATT	GGA
								~					
590			600			61	LO		(	520			630
GCT	GGG	AAG	TCT	AGC	TTT	TTC	AAC	TCA	GTG	AGG	TCT	GTT	TTC
		64	10			550			660			67	70
CAA	GGG	CAT	GTA	ACG	CAT	CAG	GCT	TTG	GTG	GGC	ACT	AAT	ACA
					•								
3.CM	ccc (	580	mcm	CAC	690	mam	3.00	70	00	mem	3 000	710	C 3 C
ACT	GGG	ATA					AGG	ACA	TAC	TCT	ATT	AGA	GAC
	720												TCA
		GAI							A11			GAC	1CA
_			_							_		٠	
7 (	60 GGG	CTC	እርጥ	770 GAG	222	GAA	780	ccc	CTC	79	90 3.TC	CAM	CAC
						_							
800 ATA			810 ATC		ልልሮ			<b>ል</b> ጥጥ		330 GAT		ጥልሮ	840 CAG
~													
			- 0			266			070				
ጥጥጥ	<b>አ</b> ልጥ							тта			САТ		TAC

		8	390	221	<b></b>	900		a.a	91	LO	0 h m	mcm	20	221
	ATT	GAT	TCC	CCA	TCG	CTG	AAG	GAC	AGA	ATT	CAT	TGT	GTG	GCA
		030			9.4	10			250			960		
	TTT	GTA	TTT	GAT	GCC	AGC	TCT	ATT	GAA	TAC	TTC	TCC	TCT	CAG
	97	70		9	980			990			100	0		
	ATG	ATA	GTA	AAG	ATC	AAA	AGA	ATT	CGA	AGG	GAG	TTG	GTA	AAC
1,	110						10	30.		10				1050
Τ/	GCT	GGT	GTG	GTA	CAT	GTG	GCT	TTG	CTC	ACT	CAT	GTG	GAT	AGC
										1080				
	ATG	GAT	CTG	ATT	ACA	AAA	GGT	GAC	CTT	ATA	GAA	ATA	GAG	AGA
		1	100	ama		1110		<b>6 7 1</b>	112	20	O.T.O.	11	130	
	TGT	GTG	CCT	GTG	AGG	TCC	AAG	CTA	GAG	GAA	GTC	CAA	AGA	AAA
		1140				E 0		٠,٠	160			1170		
	CTT	GGA	TTT	GCT	CTT	TCT	GAC	ATC	TCG	GTG	GTT	AGC	AAT	TAT
	1:	180		1:	190			1200			12	LO		
	TCC	TCT	GAG	TGG	GAG	CTG	GAC	CCT	GTA	AAG	GAT	GTT	CTA	ATT
12	220		:	1230			12	40		13	250		:	1260
	CTT	TCT	GCT	CTG	AGA	CGA	ATG	CTA		GCT				
	mm x	CAC	12	70	CCM	12	280	CAA	እጠል	1290 GGG	አአጥ	СПУ	130	00 GNG
		1 .	310			1320			13	30				
	GAA	ATT	ATC	AAC	TGT	GCA	CAA	GGA	AAA	AAA	3'			
											5'			

wherein the sign "-" represents a base complementary to the base shwon just above each sign.

- 5. An expression vector in which a DNA fragment containing a base sequence coding for a non-A non-B hepatitis-specific antigen is intoduced into a cloning site present downstream from a promoter of said vector.
- 6. The expression vector in accordance with claim 5, in which the promoter is controllable by a regulatory factor.

5

15

20

25

30

35

40

45

55

65

- 7. The expression vector in accordance with claim 5, in which the promoter operates in a microorganism.
  - 8. The expression vector in accordance with claim 5, in which the promoter operates in an eukaryote.
- 9. A transformant obtained by transforming a host with an expression vector in which a DNA fragment containing a base sequence coding for a non-A non-B hepatitis-specific antigen is introduced into a cloning site present downstream from a promoter of said vector.
- 10. The transformant in accordance with claim 9, in which the host is Escherichia coll or Bacillus subtilis.
- 11. A process for producing an antigen specific to non-A non-B hepatitis comprising introducing a DNA fragment containing a base sequence coding for said specific antigen into a cloning site present downstream from a promoter of a vector for expression, introducing the expression vector containing said DNA fragment into a host, culturing said transformed host, and collecting the produced and accumulated antigen.

17

## Fig. la

10 20 5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG 50 CAT GAA AAG ATC CTG CAA AAT CAT TTT GGA 70 GGG AAG CGG CTT AGC CTT CTC TAT AAG GGT 110 100 AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG 130 140 CTT GAC AGA TGT TGT AAT CAA GGG CCT ACT 160 170 180 CTA ACA GTG ATT TAT AGT GAA GAT CAT ATT 190 200 ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG 220 230 GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT 250 260 GCA CTT CAA GAG ACT AAA ATT TCA GAA TGG 280 290 300 AAA CTA GGA CTA TAT ACA CCA GAA ACA CTG

## Fig. 1b

310 320 TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC 340 350 CCA ACT AAT TTC CAG ATA GAT GGA AGA AAT 370 380 AGA AAA GTG ATT ATG GAC TTA AAG ACA ATG 410 400 GAA AAT CTT GGA CTT GCT CAA AAT TGT ACT 430 440 ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA 460 470 480 TGC GAA GAT TCA CTG GAC GAA AGA. AAG ATA 490 500 510 AAA GGG GTC ATT GAG CTC AGG AAG AGC TTA 520 530 540 CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT 550 560 GGA TCC CTG GTT CAA CAA ATA CGA ATT CTG 580 590 600 CTG CTG GGT CCA ATT GGA GCT GGG AAG TCT

## Fig. lc

610 620 AGC TIT TTC AAC TCA GTG AGG TCT GTT TTC 640 650 660 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG 670 680 GGC ACT AAT ACA ACT GGG ATA TCT GAG AAG 700 710 720 TAT AGG ACA TAC TCT ATT AGA GAC GGG AAA 730 740 GAT GGC AAA TAC CTG CCA TTT ATT CTG TGT 770 760 GAC TCA CTG GGG CTG AGT GAG AAA GAA GGC 790 800 GGC CTG TGC ATG GAT GAC ATA TCC TAC ATC 820 830 TTG AAC GGT AAC ATT CGT GAT AGA TAC CAG 850 860 TTT AAT CCC ATG GAA TCA ATC AAA TTA AAT 890 880 CAT CAT GAC TAC ATT GAT TCC CCA TCG CTG

## Fig. ld

910 920 930
AAG GAC AGA ATT CAT TGT GTG GCA TTT GTA

940 950 960 TTT GAT GCC AGC TCT ATT GAA TAC TTC TCC

970 980 . 990 TCT CAG ATG ATA GTA AAG ATC AAA AGA ATT

1000 1010 1020
CGA AGG GAG TTG GTA AAC GCT GGT GTG GTA

1030 1040 1050 CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC

1060 1070 1080 ATG GAT CTG ATT ACA AAA GGT GAC CTT ATA

1090 1100 1110

GAA ATA GAG AGA TGT GTG CCT GTG AGG TCC

1120 1130 1140
AAG CTA GAG GAA GTC CAA AGA AAA CTT GGA

1150 1160 1170 TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC

1180 1190 1200
AAT TAT TCC TCT GAG TGG GAG CTG GAC CCT

Fig. 2

AAAAATTTATTTGCTTTCAGGAAAATTTTTCTGT TTTTTAAATAAACGAAAGTCCTTTTAAAAAAGACA

ATAATGTGTGGAATTGTGAGCGGATAACAATTTC TATTACACACCTTAACACTCGCCTATTGTTAAAG

## Fig. le

1210 1220 1230
GTA AAG GAT GTT CTA ATT CTT TCT GCT CTG

1240 1250 1260

AGA CGA ATG CTA TGG GCT GCA GAT GAC TTC

1270 1280 1290 TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG

1300 1310 1320

<u>AAT CTA AGG GAG GAA ATT ATC AAC TGT GCA</u>

1330 CAA GGA AAA AAA 3' 5'

300 CCA Pro	360 GAA Glu	420 TGC Cys	480 CTG Leu	540 CTG Leu	600 CAA Gln
TCC	ATG	CGA Arg	TTA Leu	CTG	TTC
AAC TCC Asn Ser	ACA	TTT Phe	AGC TTA Ser Leu	ATT Ile	GTT Val
90 TAT TYr	50 AAG Lys	.10 GTT Val	70 AAG Lys	CGA Arg	101 TCT Ser
AAA Lys	350 TTA AAG	GAA Glu	AGG Arg	530 CAA ATA CGA ATT Gln Ile Arg Ile	AGG Arg
290 GCA AAA TAT A Ala Lys Tyr A	GAC	410 CAG GAT TAT GAA GTT Gln ASP TYr Glu Val	460 ATT GAG CTC AGG AAG I Ile Glu Leu Arg Lys	CAA Gln	580 AAC TCA GTG AGG TCT Asn Ser Val Arg Ser
0 GTT Val	ATG Met	OGAT ASP	GAG GAG Glu	CAA Gla	30 TCA Ser
280 GAC GTT Asp Val	34 ATT Ile	40 CAG Gln	46 ATT Ile	52 GTT Val	58 AAC Asn
TGT Cys	340 GTG ATT ATG GAC : Val Ile Met Asp ]	ATT Ile	GTC Val	520 A TCC CTG GTT CAA C ? Ser Leu Val Gln G	TTC
тст Суз	AA. Lys	TCT Ser	GGG G1y	TCC	TTT Phe
270 TTT Phe	330 AGA Arg	390 116	440 GAA AGA AAG ATA AAA Glu Arg Lys Ile Lys	35. 35. 37.	570 GGA GCT GGG AAG TCT AGC Gly Ala Gly Lys Ser Ser
CTG Leu	AAT Asn	ACT	ATA Ile	TAT Tyr	TCT Ser
ACA	AGA Arg	TGT ACT P	AAG. Lys	CCA TAT (Pro Tyr	AAG Lyb
260 CCA GAA Pro Glu	320 ATA GAT GGA Ile Asp Gly	380 CAA AAT Gln Asn	AGA AGA Arg	500 TAT GAA ( Tyr Glu	360 GGG G1y
្នឡ	GAT Asp	CAA	GAA Glu	TAT Tyr	GCT Ala
ACA Thr	ATA Ile	GCT Ala	GAC	ACT Ihr	GGA Gly
250 CTA TAT Leu Tyr	310 TTC CAG Phe Gln	370 GGA CTT Gly Leu	430 TCA CTG Ser Leu	490 TTG AGA 7 Leu Arg	550 CCA ATT Pro Ile
25 CTA Leu	31 Trc Phe	37. 66.8 61.4	TCA Ser	45 TTG Leu	CCA Pro
GGA Gly	AAT Asn	CTT	gat Asp	GCC Ala	GGT Gly
CTA Leu	ACT Thr	AAT Asn	GAA	TCT	CTG

660 TAT TYr	720 GAC ASP	780 TTG Leu	840 CAT His	900 TTT Phe	960 CGA Arg	1020 GAT AGC ATG Asp Ser Met
AAG Lys	TGT Cys	ATC Ile	AAT Asn	GTA Val	ATT Ile	AGC Ser
GAG	CTG	TAC	TTA Leu	TTT Phe	AGA ATT Arg Ile	gat Asp
650 TCT Ser	710 RATT	770 TCC Ser	830 AAA Lys	890 GCA	950 AAA Lys	10 GTG Val
650 ATA TCT Ile Ser	710 TTT ATT Phe Ile	7 ATA Ile	8 ATC Ile	890 GTG GCA TTT Val Ala Phe	950 ATC AAA 1 Ile Lys 2	1010 CAT GTG His Val
GGG G1y	CCA	770 GAC ATA TCC Asp Ile Ser	830 TCA ATC AAA Ser Ile Lys	тст Сув	AAG Lys	ACT
O ACT Thr	0 CTG Leu	0 GAT ASP	0 GAA Glu	O CAT His	0 GTA Val	O CTC Leu
64 ACA Thr	700 TAC CTG Tyr Leu	76 ATG Met	820 ATG GAA Met Glu	880 ATT CAT Ile His	940 ATA GTA Ile Val	1000 TTG CTC
640 AAT ACA ACT ASn Thr Thr	AAA Lys	760 TGC ATG GAT ( Cys Met Asp 1	CCC	aga Aeg	ATG Met	GCT Ala
ACT Thr	GGC G1y	CTG	AAT Asn	GAC	CAG	GTG Val
630 Trg Grg GGC I	690 Gat Asp	750 GGC G1y	810 Trr Phe	870 CTG AAG Leu Lys	930 TCC TCT ( Ser Ser	970 980 990 TTG GTA AAC GCT GGT GTG GTA CAT Leu Val Asn Ala Gly Val Val His
GTG Val	AAA Lys	66C 61y	CAG Gln	CTG	TCC	GTA Val
TTG	GGG Gly	GAA GGC (	TAC	860 GAT TCC CCA TCG ASP Ser Pro Ser	TTC Phe	GTG Val
. 620 CAG GCT Gln Ala	680 AGA GAC Arg Asp	740 GAG AAA Glu Lys	800 GAT AGA ASP Arg	860 CCA Pro	920 GAA TAC Glu Tyr	980 GGT G1y
CAG Gln		GAG Glu	gat Asp	TCC Ser	GAA Glu	GCT Ala
CAT His	ATT Ile	AGT Ser	CGT	gat Asp	att Ile	AAC
610 GTA ACG Val Thr	670 TAC TCT Tyr Ser	730 GGG CTG 2 Gly Leu	790 AAC ATT ASN ILE	850 TAC ATT ( Tyr Ile 2	e G	'0 GTA Val
61 GTA Val	67 TAC TYF	73 666 61y	79 AAC Asn	85 TAC	910 AGC TV Ser S	97 TTG Leu
CAT His	ACA Th <i>r</i>	CTG	GGT Gly	GAC	GCC	GAG Glu
666 61y	AGG	TCA	AAC	CAT His	gat Asp	AGG
	•					

1080	C AAG	Ser Lys	1140	AGC AAT	r Asn	1200	CTG AGA	u Arg	1260	G AAT	Gly Asn			
	Ę	Se		AG	Se		ີ້ວ	Le		Ö	G			
	AGG	Arg		GIT	Val		ပ္ပ	Ala	1250	ATA	Ile			
070	GTG	Pro Val	1130	GTG	Val	1190	-	Ser	250	CAA	Glu			
-	ຽ	Pro	-	TCG	Ser	٦	CTI	Leu	-	GAG	Glu			
	GTG	Val	1120	ATC	Ile		H	]e		TIT	Phe			
09	TGT	Glu Arg Cys	20	GAC	Asp	80	GAT GTT CTA A	Leu	40	CCI	Leu Pro	00	AAA AAA TAG	*
10	AGA	Arg	11	TCI	Ser	11	GIT	Val	12	TIG	Leu	13	AAA	Lys
	GAG	Glu		r L L	Leu		GAT	Asp		GAT	Glu Asp I		AAA	Lys
	ATA	Ile		GCF	Ala		AAG	Гұз		GAG	Glu		GGA	313
1050	GAA	Gly Asp Leu Ile Glu	1110	TII	Phe	1170	GTA	Glu Leu Asp Pro Val	1230	TTA	Ala Asp Asp Phe Leu	.290	GCA CAA	Gln
•	ATA	Ile	•	<b>S</b>	Gly	_	CCI	Pro	_	TIC	Phe		<b>₹</b>	Ala
	CTT	Leu		E C	Leu		GAC	Asp		GAC	Asp		ATC AAC TGT	Сув
040	GAC	Asp	1100	AA	Lys	091	CIG	Leu	120	GAT	Asp	80	AAC	Asn
Ä	GGT	Gly	F	AGA	Arg	7	GAG	G1u	12	SCA SCA	Ala	12	ATC	Ile
1030	A	Lys		S C	GIn		TGG	Trp	1210	CCI	Ala	1270	ATT	Ile
20	ACA	Thr	1090	GIC	Val	00	GAG	GIu	9	TGG	Trp	0	GAA	Gla
10	ATT	Ile	100	GAA	Glu	118	TCT	Ser .	12]	CIA	Leu	127	GAG	Gla
	STS	Leu		GAG			TCC TCT GAG	Ser		ATG	Met		AGG	Æg
	GAT	Asp		CIA	Leu		TAT	Tyr		CGA	Arg		CIA	Leu

CTATTCAGGA TTATGAAGTT TTTCGATGCG AAGATTCACT GGACGAAAGA AAGATAAAAG GGGTCATTGA GCTCAGGAAG GATAAGTCCT AATACTTCAA AAAGCTACGC TTCTAAGTGA CCTGCTTTCT TTCTATTTTC CCCAGTAACT CGAGTCCTTC

80	150	240	320	400	480
TCGTTTGACA	AAGGGTAGTG TCCATGGATT	CATATTATTG	TAAAATTTCA	CTAATTTCCA	TGTACTATCT
AGCAAACTGT	TTCCCATCAC AGGTACCTAA	GTATAATAAC	ATTTTAAAGT	GATTAAAGGT	ACATGATAGA
70	150	230	310	390	470
CAGTGACAAC	AAGGGTAGTG	TAGTGAAGAT	TTCAAGAGAC	AACTCCCCAA	TGCTCAAAAT TGTACTATCT
GTCACTGTTG	TTCCCATCAC	ATCACTTCTA	AAGTTCTCTG	TTGAGGGGTT	ACGAGTTTTA ACATGATAGA
50	140	220	300	380	460
CAACAGATCA AGAAGTATGG	CCTTCTCTAT	CAGTGATTTA	CTTTTTGCAC	TGCAAAATAT	ATCTTGGACT
GTTGTCTAGT TCTTCATACC	GGAAGAGATA	GTCACTAAAT	GAAAAACGTG	ACGTTTTATA	TAGAACCTGA
	120 170 TTTGGAGGGA AGCGGCTTAG AAACCTCCCT TCGCCGAATC	210 CCTACTCTAA GGATGAGATT	290 TTCCATCATC AAGGTAGTAG	370 GTTGTGACGT CAACACTGCA	450 ACAATGGAAA TGTTACCTTT
40 ACAGACAGTA TGTCTGTCAT		200 TAATCAAGGG ATTAGTTCCC		350 TACACCAGAA ACACTGTTTT ATGTGGTCTT TGTGACAAAA	440 GGACTTAAAG CCTGAATTTC
30	110	190	60 270 280	350	430
AGCTCATACT	GCAAAATCAT	ACAGATGTTG	GT TACCAGGAAA GAAAGTATGC	TACACCAGAA	AAGTGATTAT
TCGAGTATGA	CGTTTTAGTA	TGTCTACAAC	CA ATGGTCCTTT CTTTCATACG	ATGTGGTCTT	TTCACTAATA
20	100	170	260	340	410
CCTCAGCTCT	AAAAGATCCT	CCATAATGGA GTTTTGCTTG	AGAAGAGGT	GAATGGAAAC TAGGACTATA	GATAGATGGA AGAAATAGAA
GGAGTCGAGA	TTTTCTAGGA	GGTATTACCT CAAAACGAAC	TCTTCCCCA	CTTACCTTTG ATCCTGATAT	CTATCTACCT TCTTTATCTT
10	90	170	250	330	410
GGGGGCTAC CCTCAGCT	TGGTTGCATG AAAAGATC	CCATAATGGA	GAGCATATGC AGAAGAGG	Gaatggaaac	GATAGATGGA
CCCCCGATG GGAGTCGA	ACCAACGTAC TTTTCTAG	GGTATTACCT	CTCGTATACG TCTTCTCC	Cttacctytg	CTATCTACCT
ສີຄ					

					•	
640	720	800	880	960	1040	1120
TGGGTCCAAT	TTGGTGGGCA	GCCATTTATT	ACGGTAACAT	TCGCTGAAGG	AAAGATCAAA	ATCTGATTAC
ACCCAGGTTA	AACCACCGT	CGGTAAATAA	TGCCATTGTA	AGCGACTTCC	TTTCTAGTTT	TAGACTAATG
630	710	790	870	950	1030	1110
ATTCTGCTGC	GCATCAGGCT	GCAAATACCT	TACATCTTGA	TGATTCCCCA	AGATGATAGT	GATAGCATGG
TAAGACGACG	CGTAGTCCGA	CGTTTATGGA	ATGTAGAACT	ACTAAGGGGT	TCTACTATCA	CTATCGTACC
620	700	780	860	940	1020	1100
ACAAATACGA	GGCATGTAAC	GGGAAAGATG	TGACATATCC	ATGACTACAT	TTCTCCTCTC	CACTCATGTG
TGTTTATGCT	CCGTACAITG	CCCTTTCTAC	ACTGTATAGG	TACTGATGTA	AAGAGGAGAG	GTGAGTACAC
610	690	770	850	930	1010	1090
CCCTGGTTCA	GTTTTCCAAG	TATTAGAGAC	TGTGCATGGA	TTAAATCATC	TATTGAATAC	TGGCTTTGCT
GGGACCAAGT	CAAAAGGTTC	ATAATCTCTG	ACACGTACCT	AATTTAGTAG	ATAACTTATG	ACCGAAACGA
600	680	760	840	920	1000	1080
CCATATGGAT	AGTGAGGTCT	GGACATACTC	GAAGGCGGCC	ATCAATCAAA	TTTGTATTTG ATGCCAGCTC	GTGGTACATG
GGTATACCTA	TCACTCCAGA	CCTGTATGAG	CTTCCGCCGG	TAGTTAGTTT	AAACATAAAC TACGGTCGAG	CACCATGTAC
590	670	750	830	910	990	1070
AACTTATGAA	TTTTCAACTC	GAGAAGTATA	GAGTGAGAAA	ATCCCATGGA	TTTGTATTTG	AAACGCTGGT
TTGAATACTT	AAAAGTTGAG	CTCTTCATAT	CTCACTCTTT	TAGGGTACCT	AAACATAAAC	TTTGCGACCA
580	660	730	810	900	980	1050
CTGCCTTGAG	AAGTCTAGCT	CTAATACAAC TGGGATATCT	CTGTGTGACT CACTGGGGCT	TCGTGATAGA TACCAGTTTA	ACAGAATTCA TTGTGTGGCA	AGAATTCGAA GGGAGTTGGT
GACGGAACTC	TTCAGATCGA	GATTATGTTG ACCCTATAGA	GACACACTGA GTGACCCCGA	AGCACTATCT ATGGTCAAAT	TGTCTTAAGT AACACACGT	TCTTAAGCTT CCTCAACCA
570	650	730	810	890	970	1050
AGCTTACTGT	TGGAGCTGGG	CTAATACAAC	CTGTGTGACT	TCGTGATAGA	ACAGAATTCA	AGAATTCGAA
TCGAATGACA	ACCTCGACCC	GATTATGTTG	GACACACTGA	AGCACTATCT	TGTCTTAAGT	TCTTAAGCTT

1610 1620 1630 1640 1650 CTTGGATAAA ATAAATTTCT TATAAAAAA AAAAAAAA AAAAAAA 3'GAACCTAAAT ACAAGACATA GACACTTTTT TATTTAAAGA ATATTTTTTT TTTTTTTT TTTTTTT TTTTTTT TT

1200	1280	1360	1440	1520	1600
CTTGGATTTG	AATTCTTTCT	TAAGGGAGGA	AGATTAAAAT	GATGAAGAAA	AATAATTTTT
GAACCTAAAC	TTAAGAAAGA	ATTCCCTCCT	TCTAATTTTA	CTACTTT	TTATTAAAAA
1190	1270	1350	1430	1510	1590
CCAAAGAAAA	AGGATGTTCT	ATAGGGAATC	ACATCACAGA	TAATGTCTAG	GAAAATAAT
GGTTTCTTTT	TCCTACAAGA	TATCCCTTAG	TGTAGTGTCT	ATTACAGATC	CTTTTTATTA
1180	1260	1340	1420	1500	1580
TAGAGGAAGT	GACCCTGTAA	TTTTGAGCAA	AAATTTCCTC	TGTGTTTAT	TCATAATTGT
ATCTCCTTCA	CTGGGACATT	AAAACTCGTT	TTTAAAGGAG	ACACAAAATA	AGTATTAACA
1170	1240	1320	1410	1480	1560
AGGTCCAAGC	ATTCCTCTGA GTGGGAGCTG	GACTTCTTAG AGGATTTGCC	AGGTTCACGT	AGTAACTAAG ACCAAAGGA	CTAGAAATAA CATGATTTAG
TCCAGGTTCG	TAAGGAGACT CACCCTCGAC	CTGAAGAATC TCCTAAACGG	TCCAAGTGCA	TCATTGATTC TGGTTTCCCT	GATCTTTATT GTACTAAATC
1160	1240	1320	1400	1480	1560
TGTGCCTGTG	ATTCCTCTGA	GACTTCTTAG	GATATGTGAA	AGTAACTAAG	Ctagaaataa
ACACGGACAC	TAAGGAGACT	CTGAAGAATC	CTATACACTT	TCATTGATTC	Gatctttatt
1150	1230	1310	1390	1470	1550
TAGAGAGATG	GTTAGCAATT	GGCTGCAGAT	Gaaaaaaata	ACCAAAGAGA	TTGTAAATAA
ATCTCTCTAC	CAATCGTTAA	CCGACGTCTA	Cittititat	TGGTTTCTCT	AACATTTATT
1130	1210	1290	1370	1460	1530
AAAAGGTGAC CTTATAGAAA	CTCTTTCTGA CATCTCGGTG	GCTCTGAGAC GAATGCTATG	AATTATCAAC TGTGCACAAG	Gaaaacacag	TGCATAGAAC ATTGTAGTAC
TTTTCCACTG GAATATCTTT	GAGAAAGACT GTAGAGCCAC	CGAGACTCTG CTTACGATAC	TTAATAGTTG ACACGTGTTC	Cttttggggg	ACGTATCTTG TAACATCATG
1130	1210	1290	1370	1450	1530
AAAAGGTGAC	CTCTTTCTGA	GCTCTGAGAC	AATTATCAAC	TCAGAAAGGA	TGCATAGAAC
TTTTCCACTG	GAGAAAGACT	CGAGACTCTG	TTAATAGTTG	AGTCTTTCCT	ACGTATCTTG
	•				

Fig.5

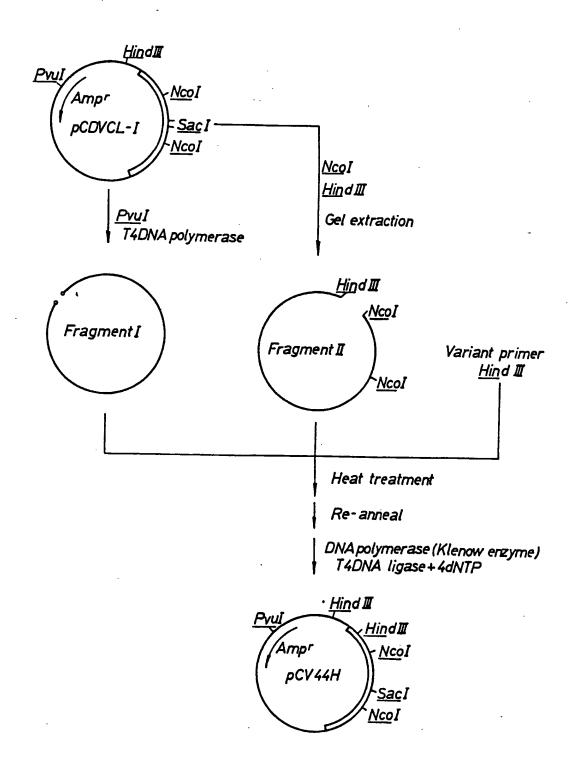


Fig.6

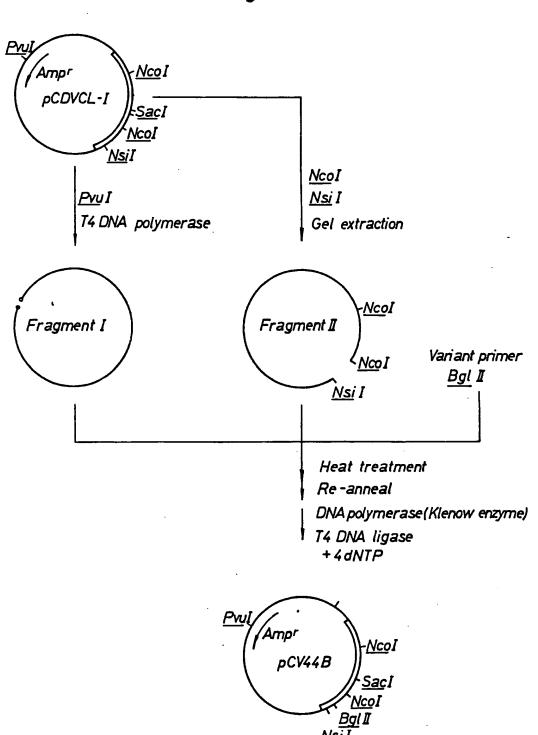
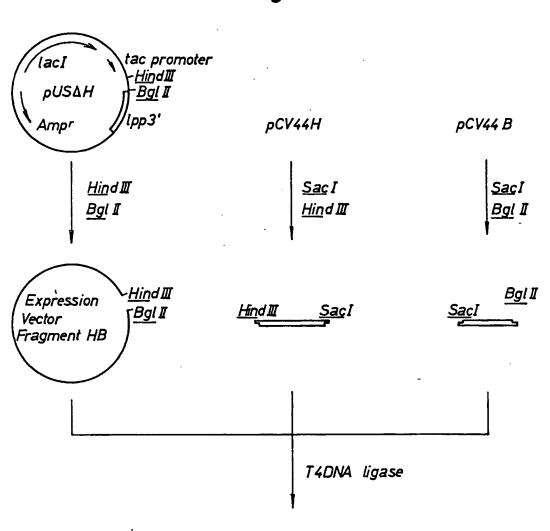
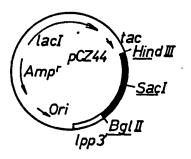


Fig.7







# EUROPEAN SEARCH REPORT

Application Number

EP 88 40 0790

	<del></del>			EP 88 40 07
		IDERED TO BE RELEVA	NT	
Category	Citation of document with of relevant p	indication, where appropriate,	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl. 4)
D,Y	EP-A-0 190 972 (M * Whole document *	ITSUBISHI)	1-11	C 12 N 15/00 A 61 K 39/29
D,Y	PROC. NATL. ACAD. S March 1983, pages et al.: "Efficient by using antibody particle *	1194-1198; R.A. YOUNG isolation of genes	1-11	, , oz k
	"Non-A, non-B hepat Ag is a normal cell	, vol. 105, no. 17, 527, no. 151110y, T. AKATSUKA et al.: citis related AN6520 Jular protein mainly II" & J. MED. VIROL.	1	
	& J. MED. VIROL 15(	MATSU et al.: antigen purified I-A. non-B hepatitis"	1	TECHNICAL FIELDS SEARCHED (Int. CL4)
- 1	* Whole abstract *`  EP-A-0 066 296 (EI	SAI CO., LTD)		C 12 N A 61 K G 01 N
	EP-A-0 092 249 (EI	•		
	The present search report has b	een drawn up for all claims		
	Place of search	Date of completion of the search	<del></del>	Examiner
THE	HAGUE	21-06-1988	SKELI	LY J.M.
X : partic Y : partic docui A : techn	ATEGORY OF CITED DOCUMES  cularly relevant if taken alone  cularly relevant if combined with and ment of the same category  ological background  written disclosure	E : earlier patent of after the filing ther D : document cited L : document cited	in the application	bed on, or

men was made



Europäisches Patentamt

**European Patent Office** 

Office européen des brevets



(1) Publication number: 0 293 274 B1

(12)

## **EUROPEAN PATENT SPECIFICATION**

(4) Date of publication of patent specification: 04.09.91 Bulletin 91/36

(61) Int. CI.5: C12N 15/12, C12N 15/33

(21) Application number: 88400790.7

2 Date of filing: 31.03.88

- 64) DNA molecules encoding non-A, non-B hepatitis antigens, and their use in producing said antigens.
- (30) Priority: 31.03.87 JP 78313/87 04.06.87 JP 140586/87 10.11.87 JP 283990/87
- (3) Date of publication of application: 30.11.88 Bulletin 88/48
- (48) Publication of the grant of the patent: 04.09.91 Builetin 91/36
- Designated Contracting States:
   CH DE FR GB IT LI NL SE
- (36) References cited:
  EP-A- 0 068 298
  EP-A- 0 092 249
  EP-A- 0 190 972
  PROC. NATL. ACAD. SCI. USA, vol. 80, March 1983, pages 1194-1198; R.A. YOUNG et al.:
  "Efficient isolation of genes by using antibody probes"
  CHEMICAL ABSTRACTS, vol. 105, no. 17, October 1986, page 527, no. 151110y, Columbus, Ohio, US; T. AKATSUKA et al.: "Non-A, non-B hepatitis related AN6520 Ag is a normal cellular protein mainly expressed in liver ii" & J. MED. VIROL. 1986, 20(1), 43-56
  BIOLOGICAL ABSTRACTS, vol. 80, 1985, no. 4871; J.-I. TOHMATSU et al.: "AN6520 antigen: an antigen purified from liver with non-A, non-B hepatitis" & J. MED. VIROL. 15(4): 357-372, 1985.

- (3) Proprietor: MITSUBISHI KASEI CORPORATION 5-2, Marunouchi 2-chome Chiyoda-ku Tokyo 100 (JP)
- (72) Inventor: Takahashi, Kazuhiro 303 Corpo **Omodaka** 29 Ichijoji-Somedono-cho Sakyo-ku Kyoto City Kyoto Prefecture (JP) Inventor: Shibul, Tatsurou M606-11 San White 6-15-4 Minami-Naruse Machida City Tokyo (JP) Inventor: Kamizono, Michiru F202-24 San White 2-8-11 Tsuruma Machida City Tokyo (JP) Inventor: Matsul, Rie A211-23 San White 1-26-31 Ogawa Machida City Tokyo (JP) Inventor: Teranishi, Yutaka 1-14-2 Takane Sagamihara City Kanagawa Prefecture (JP) Inventor: Nakanishi, Shigetada 517-116 Iwakura-Nagatani-cho Sakyo-ku Kyoto City Kyoto Prefecture (JP) Inventor: Kitamura, Naomi 2-272 Yakumo-Higashi-machi Moriguchi City Osaka Prefecture (JP)
- (2) Representative: Gutmann, Ernest et al S.C. Ernest Gutmann - Yves Plasseraud 67, boulevard Haussmann F-75008 Paris (FR)

EP 0 293 274 B

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

#### EP 0 293 274 B1

#### Description

10

The present invention generally relates to the production of an antigen specific to non-A non-B hepatitis by recombinant DNA technology. More particularly, it relates to a DNA fragment coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis, an expression vector containing such a DNA fragment, a host transformed with such an expression vector, as well as a process for producing said antigen specific to non-A non-B hepatitis by culturing such a transformant.

Among viral hepatitises, the viral entities of hepatitis type A and type B have been found and, accordingly, it has now become possible to diagnose such diseases by immunological methods.

Still another type of hepatitis different from the types A and B, which is called non-A non-B type hepatitis, is said to be over 90% of post-transfusion hepatitis: refer to NIPPON RINSHO (Japan Clinic), 35, 2724 (1977); J. Biol. Med., 49, 243 (1976). The pathogenic virus of the non-A non-B type hepatitis, however, has not yet been identified. Only one fact which has already been established is potential infection of human hepatitis type non-A non-B virus to chimpanzee: refer to Lancet I, 459 (1978); ibid., 463 (1978).

Many workers have done various investigations for searching an antigen-antibody system related to the non-A non-B hepatitis by using mainly sera from patients affected with the disease; nevertheless, no definite system has been found. Under these circumstances, the diagnosis of non-A non-B hepatitis should inevitably be effected by so-called exclusion diagnosis: that is, whether or not the hepatitis of a patient is type A or type B or other hepatitis due to a virus known to cause hepatopathy, for example, CMV, HSV, EBV, etc., is first determined; and if not, the patient's hepatitis is diagnosed as non-A non-B type. Thus, such a diagnosis of non-A non-B hepatitis will require much time and labor.

An antigenic protein specific to non-A non-B hepatitis and useful for the direct diagnoses of the hepatitis has been purified from human and chimpanzee hepatocytes affected with non-A non-B hepatitis, and a monoclonal antibody specific to the antigen and useful for the treatment of the non-A non-B hepatitis has also been proposed: refer to Japanese Patent Application Laying-open (KOKAI) Nos. 176856/86 and 56196/86.

A large amount of such an antigenic protein specific to non-A non-B hepatitis should be required when such a protein is to be employed, for example, as a diagnostic agent. However, it is not always appropriate to purify such a large amount of the antigenic protein from chimpanzee hepatocytes affected with non-A non-B hepatitis.

On the other hand, in order to detect a gene coding for a specific antigen of non-A non-B hepatitis by nucleic acid hybridization and, further, to produce such an antigen specific to non-A non-B hepatitis by the recombinant DNA technology, it is essential to obtain a gene fragment coding for the antigenic protein specific to the non-A non-B hepatitis.

The present inventors have made great efforts to produce such a specific antigenic protein in a large amount by genetic engineering techniques, and finally isolated a gene fragment coding for the antigenic protein specific to non-A non-B hepatitis, said gene fragment being useful for the production of such antigens. Further, the inventors have successfully constructed an expression vector containing said gene fragment. Thus, the present invention has now been attained.

It is an object of the invention to provide a DNA fragment which contains a base sequence coding for an antigen specifically occurring in a host cell affected with non-A non-B hepatitis or an antigenic protein specific to non-A non-B hepatitis having physiological activities equivalent to those of said specifically occurring antigen.

Another object of the invention is to provide an expression vector having said DNA fragment introduced thereinto at a cloning site downstream from a promoter of the vector.

A still another object of the invention is to provide a transformant obtained by transforming a host cell with said expression vector.

A further object of the invention is to provide a process for producing such an antigen specific to non-A non-B hepatitis by culturing said transformant.

Other objects and advantages of the present invention will be apparent from the following detailed description with reference to the attached drawings, in which:

Figs. 1a-1e show the base sequence coding for an antigenic protein specific to non-A non-B hepatitis; Fig. 2 shows the base sequence of a hybrid promoter Pac;

Figs. 3a-3c show the base sequence of a cDNA fragment obtained in Example 1 described hereinbelow, together with deduced amino acid sequence;

Figs. 4a-4c show the base sequence of cDNA containing the full length gene sequence of an antigenic protein specific to non-A non-B hepatitis, which cDNA was obtained in Example 2 described hereinbelow, the base sequence 57-1388 thereof coding for the antigenic protein specific to non-A non-B hepatitis;

Fig. 5 schematically illustrates the construction of a plasmid pCV44H:

Fig. 6 schematically illustrates the construction of a plasmid pCV44B; and

Fig. 7 schematically illustrates the construction of a plasmid pCZ44.

#### EP 0 293 274 B1

The present invention will be described in detail hereinbelow.

10

20

According to one aspect of the invention, a DNA fragment is provided which contains a base sequence coding for an antigenic protein occurring specifically in hepatocytes affected with non-A non-B hepatitis.

Such a DNA fragment of the invention may be prepared in the following manner.

First, a liver tissue specimen derived from a human or chimpanzee individual affected with non-A non-B hepatitis is homogenized in an aqueous solution of guanidinium thiocyanate and then subjected to cesium chloride equilibrium density gradient centrifugation according to Chirgwin et al. method (Biochemistry, 18, 5294-5299 (1979)) to separate total RNA as a precipitate. After separation, the total RNA is purified by phenol extraction and ethanol precipitation.

"Individuals affected with non-A non-B hepatitis" used as sources of liver tissue specimens in the invention may include those affected with so-called type D hepatitis, which has recently been named.

It is known that mRNA of an antigen gene generally has a poly-A chain. Thus, the total RNA is subjected to oligo(dT) cellulose column chromatography in a conventional manner and poly(A)-containing RNA (poly A\* RNA) is isolated as mRNA material.

A cDNA library corresponding to the poly A+ RNA is then obtained from the mRNA material according to the random primer method (Y. Ebina et al., Cell, 40, 747-758 (1980)): Thus, a number of DNAs complementary to the mRNA material are randomly synthesized using any primer of e.g. about 6 bases and a reverse transcriptase.

The cDNA is methylated with a DNA methylase, e.g. EcoRI methylase, to protect cleavage sites present in the cDNA capable of being cleaved by a corresponding restriction enzyme, e.g. EcoRI. A DNA linker containing the corresponding restriction enzyme cleavage sites at both ends, e.g. EcoRI linker (CGAATTCG), is added to the methylated cDNA and, then, this cDNA is digested with the restriction enzyme, e.g. EcoRI.

The digested cDNA is then cloned into a cloning vector such as a plasmid or a λ phage. For example, the cDNA may be introduced into EcoRI site of λgt 11 DNA, which is an expression cloning vector: refer to R.A. Young et al., Pro. Nati. Acad. Sci. U.S.A., 80, 1194-1198 (1983). The cDNA will be inserted into the β-gal gene on the λgt 11 phage. Thus, expression of the cDNA can be easily verified by the production of a fused protein with β-galactosidase due to induction of the expression by the lactose operon promoter of said phage when Ecoli transfected with said phage is cultured in a medium containing iPTG (isopropylithio-β-D-galactopyranoside).

The  $\lambda$ gt 11 phage incorporating the cDNA is then introduced into <u>E. coll.</u> by Tomizawa <u>et al.</u> method in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). The thus transfected microorganism is cultured in an IPTG-containing medium.

The thus formed plaques can be easily selected by an immunological screening method using a monoclonal antibody specifically directed to non-A non-B hepatitis to obtain a desired cDNA. Such a monoclonal antibody which can be used in the immunological screening method may be prepared according to the methods described in Japanese Patent Application Laying-open Nos. 56196/86 and 91328/88. The screening methods used may include the western blotting technique described in these applications.

The plaques positive in the immunological screening test are selected to proliferate the phage by Tombzawa et al. method. DNA is purified from the grown phage by T. Maniatis et al. method in "Molecular Cloning", Cold Spring Harbor Laboratory, pp. 85 et seq. (1982), and cleaved with a suitable restriction enzyme such as EcoRi. The thus purified and digested DNA fragments can be used to determine the base sequence of a desired cDNA segment according to Maxam and Gilbert method in Methods in Enzymology, 65, 499-560 (1980); or alternatively, after further cloning the DNA fragments into M13 phage, the base sequence of such a desired cDNA segment can be determined according to the dideoxy method: Sanger et al., Proc. Natl. Acad. Sci. U.S.A., 74, 5463 (1977).

Thus, a cDNA fragment coding for an antigen specific to non-A non-B hepatitis can be obtained. However, such a DNA fragment may usually be only a portion of the gene coding for the non-A non-B hepatitis-specific antigen.

A full length cDNA coding for such a non-A non-B hepatitis-specific antigen may be obtained in the following manner.

Poly A\*-mRNA is isolated and purified in a manner similar to that described above. From the poly A\*-mRNA a cDNA library is obtained according to Okayama-Berg vector-primer method: Molecular and Cellular Biology, 2, 161-170 (1982).

A plasmid containing such a cDNA thus prepared is used to transform <u>E. coii</u> by any conventional method, for instance, the method D. Hanahan: J. Mol. Biol., 168, 557 (1983). The transformant ampicilin-resistant strains are collected and screened by the colony hybridization method using the aforementioned DNA fragment as a probe. Such a probe may preferably be prepared by either the strepto-avidin method, or the nick translation method using photobiotinnucleic acids and \*2P-nucleic acids.

The thus selected colonles containing a cDNA clone are cultured. Plasmid DNA is obtained from the cul-

tured colony according to Birmboim et al. method (Nucleic Acid Res., 7, 1513 (1979)) and digested with a suitable restriction enzyme. The base sequence of a desired full-length cDNA segment is then determined according to the aforementioned Maxam and Gilbert method or, alternatively, after further cloning the digested DNA into M13 phage or pVC12 plasmid, such a base sequence is determined according to the above described Sanger et al. dideoxy method.

The base sequence of the full length DNA coding for an antigen specific to non-A non-B hepatitis is shown in Fig. 1, in which the symbol "--" just under the base sequence represents a corresponding base complementary to the respective base described just above each of the symbols.

Of course, DNA fragments which can be employed in the invention do not necessarily contain the same base sequence as shown in Fig. 1, but those DNA fragments in which a part of said base sequence shown in Fig. 1 has been substituted by at least one different base or deleted therefrom and those DNA fragments in which one or more additional bases have been added to the base sequence of Fig. 1 may also be included herein provided that such different DNA fragments may code for substances having physiological activities equivalent to those of the non-A non-B hepatitis-specific antigens encoded by the base sequence of Fig. 1.

According to another aspect of the invention, an expression vector is provided in which the aforementioned DNA fragment of the invention is inserted into a cloning site downstream from a promoter of this vector.

The expression vector of the invention contains a promoter in a position capable of controlling the transcription of a DNA fragment coding for a non-A non-B hepatitis-specific antigen obtained by the aforementioned method. The promoters used in the invention may be any promoter capable of expressing the DNA fragment in a host, and preferably of controlling the transcription of the fragment.

When a host used is a microorganism such as <u>Escherichia coli</u>, <u>Bacilius subtilis</u>, etc., the expression vector of the invention may preferably comprise a promoter, a ribosome binding sequence, a gene for a non-A non-B hepatitis-specific antigen, a transcription termination factor, and a gene controlling the promoter.

The promoter used may include those derived from E. coli, phage, etc., for example, tryptophan synthase operon (trp), lactose operon (lac), lipoprotein (lpp), recA, lambda phage P<sub>L</sub>, P<sub>R</sub>. T5 early gene P<sub>25</sub>, P<sub>26</sub> promoter, which may also be prepared by chemical synthesis. Also included herein are hybrid promoters such as tac (trp: lac), trc (trp: lac) and Pac (phage: E. coli) shown in Fig. 2.

The ribosome binding sequence may be derived from E. coli, phage, etc., but preferably may be those synthetically prepared, for example, those containing a consensus sequence such as

# AGGAGGTTTAA. SD sequence

The gene for a non-A non-B hepatitis-specific antigen may be directly employed without any modification. Preferably, an unnecessary base sequence (non-coding region) may be deleted by site-directed mutagenesis: BIO TECHNOLOGY, July, 636-639 (1984).

A transcription termination factor may not always be required in the expression vector of the invention. Preferably, the instant vector may contain a  $\rho$ -independent terminator, for example, lpp terminator, trp operon terminator, ribosomal RNA gene terminator, etc.

The expression vector may be derived from any conventional plasmid. Preferably, it may be derived from such a plasmid as replicating itself in <u>E. coli</u> or <u>Bacillus subtilis</u>, for example, pBR322- or pUB110-derived plasmid.

Desirably, these factors required for expression are arranged in the expression plasmid in the order of the promoter, the SD sequence, the structural gene of a non-A non-B hepatitis-specific antigen, and the transcription termination factor from 5' to 3'. A repressor gene required to control the transcription, a marker gene such as drug-resistant gene, and a plasmid replication origin may be arranged in any order in the expression vector.

The expression vector of the invention may be introduced into a host by any conventional method for transformation of E. coli, e.g., one described in Molecular Cloning, 250-253 (1982), or of Bacilius subtilis, e.g., one described in Molec. Gen. Genet., 168, 111-115 (1979) or Proc. Nat. Acad. Sci. U.S.A., 44, 1072-1078 (1958).

The resulting transformant may be cultured in any conventional medium, e.g. one described in Molecular Cloning, 68-73, (1972), at a temperature in the range of 28 to 42°C in both cases of E. coli and Bacillus subtilis. Preferably, it may be cultured at a temperature in the range of 28 to 30°C where no expression of heat shock proteins may be induced.

The desired protein thus produced may be easily purified from the host in conventional procedures. For example, the host cell may be crushed by lysozyme-surfactant or ultra-sonication, and the insoluble fractions which contain the desired non-A non-B hepatitis-specific antigen may be then collected by centrifugation, solubilized in a surfactant such as 0.01% SDS, and subjected to column chromatography using a monocional anti-

body (Japanese Patent Application Laying-open (KOKAI) Nos. 56196/86 and 176856/88.

When an eukaryotic cell such as an animal cell is employed as a host, the expression vector of the invention is preferably as follows:

The promoters used in the vector of the invention for the expression in eukaryotic cells may herein include SV40 early and late promoters; promoters of apolipoprotein E and A-I genes; promoter of heat shock protein gene (Proc. Natl. Acad. Sci. U.S.A., 78, 7038-7042 (1981)); promoter of metallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6511-6515 (1980)); HSV TK promoter; adenovirus promoter, such as Ad2 major late promoter (Ad2 MLP); LTR (long terminal repeat) of retrovirus; etc. SV40 promoter and promoter of metallothionein gene are preferred.

The expression vector of the invention may contain a splice sequence comprising 5' splice junction donor site, an intron and 3' splice junction acceptor site. A common base sequence is found at all the splice junction sites (exonintron junction sites); so-called GT/AG rule that any intron region always starts from two bases GT at the donor site and terminates at two bases AG of the acceptor site has been established.

The expression vector of the invention may contain one or more splice sequences as mentioned just above. Such splice sequences may be positioned upstream or downstream of the structural gene for a non-A non-B hepatitis-specific antigen.

illustrative examples of such splice sequences may include those DNA sequences found in exons 2 and 3 of rabbit β-globin gene (Science, 26, 339 (1979)) and mouse methallothionein-I gene containing the promoter, exons 1, 2 and 3 and introns A and B of methallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6513 (1980)). The 5' and 3' splice sites may be derived from the same or different gene; for example, a sequence in which 5' splice site contained in adenovirus DNA is linked to 3' splice site derived from the gene of Ig variable region can be employed.

The expression vector of the invention also contains a polyadenylation site downstream from the structural gene of a non-A non-B hepatitis-specific antigen. Illustrative examples of the polyadenylation sites may include those derived from SV40 DNA, β-globin gene or methaliothionein gene. A combined site of the polyadenylation sites of β-globin gene and SV40 DNA may be employed in the invention.

The expression vector of the invention may also contain a dominant selective marker permitting the selection of transformants. Selective markers which can be used herein may include DHFR gene imparting MTX (methotrexate) resistance to a host; tk gene of herpes simplex virus (HSV) which permits selection of tk strains transformed therewith in HAT medium; the gene for aminoglycoside 3'-phosphotransferase from E. coli transposon Tn5, which imparts to a host the resistance against 3'-deoxystreptamine antibiotic G418; bovine papilloma virus gene permitting morphological discrimination by piled up growth; and aprt gene.

Alternatively, animal cells transformed with the expression vector of the invention may be selected by the cotransformation even though no selective marker is present in the vector. For this purpose, an animal cell is cotransformed with both the expression vector and a plasmid or other DNA containing a gene for such a selective marker and selected by a phenotypic trait of the gene.

Advantageously, the expression vectors may also contain a plasmid fragment having an origin of replication derived from a bacterium such as <u>E. coli</u>, since such vectors can be cloned in bacteria. Such plasmids may include pBR322, pBR327, pML, etc.

tilustrative examples of plasmid vectors used as sources of the expression vectors according to the invention may include pKCR (Proc. Natl. Acad. Sci. U.S.A., 78, 1528 (1981)), which contains SV40 early promoter, the splice sequence and polyadenylation site derived from rabbit β-globin gene, the polyadenylation alte from SV40 early region, and the origin of replication and ampicillin resistant gene from pBR322; pKCR H2 (Nature, 307, 605 (1984)), in which the pBR322 portion of pKCR has been substituted by pBR327 fragment and the EcoRI site present in the exon 3 of rabbit β-globin gene has been converted into Hindlii site; and pBPV MT1 containing BPV gene and methallothlonein gene (Proc. Natl. Acad. Sci. U.S.A., 80, 398 (1983)).

Animal cells transformed with the expression vector of the invention may include CHO cells, COS cells, and mouse L cells, C127 cells and FM3A cells.

The introduction of the expression vector of the invention into an animal cell may be carried out by transfection, microinjection, etc. Most often, the transfection may employ CaPO<sub>4</sub>: Virology, <u>52</u>, 458-467 (1973).

Animal cells transformed by introducing the expression vector of the invention may be cultured in a suspension or solid medium by conventional methods. The culture medium used is most often MEM, RPMI1640, etc.

Proteins produced in the transformed animal cells can be separated and purified in the almost same manner as in the case of microorganisms aforementioned.

As stated, the invention provides a transformant cell obtained by introducing the expression vector of the invention into a host cell.

Also provided according to the invention is a process for producing a non-A non-B hepatitis-specific antigen

comprising culturing said transformant and collecting the produced and accumulated antigen.

As stated previously, a large amount of an antigenic protein specific to non-A non-B hepatitis is required when such a protein is to be utilized as a direct diagnostic agent. According to the present invention, such an antigenic protein can be produced with a low cost and a large scale without use of infected chimpanzee hepatocytes. Prior to the present invention, it has been difficult obtain such a large amount of a non-A non-B hepatitis-specific antigenic protein from hepatocytes of chimpanzees affected with non-A non-B hepatitis.

Further, the DNA fragment coding for an antigenic protein of non-A non-B hepatitis virus according to the present invention will be useful as a probe for detecting the gene of said antigenic protein by nucleic acid hybridization.

#### EXAMPLES

10

15

20

The following examples will be given by way of illustration but these examples in no way limit the scope of the invention without departing the concept thereof.

EXAMPLE 1 : Preparation of cDNA Fragment Coding for Antigenic Protein Specific to Non-A Non-B Hepatitis

Poly(A)-containing RNA was prepared from chimpanzee liver according to the guanidine thiocyanate-lithium chloride method: Cathala et al., DNA, 2, 329 (1983).

The infected liver (5 g) was taken out from a chimpanzee affected with non-A non-B hepatitis and immediately frozen by liquid nitrogen. The frozen liver was added into a Waring blender together with liquid nitrogen and ground at 3,000 rpm for 2 minutes. The ground liver specimen was further ground by a Teflon homogenizer at 5 rpm in 100 ml of a solution : 5 M guanidine thiocyanate, 10 mM EDTA, 50 mM Tris-HCl (pH 7), 8% (v/v) β-mercaptoethanol. The thus solubilized material (20 ml) was slowly placed on 5.7 M CsCl solution (10 ml) contained in a centrifuge tube and centrifuged at 27,000 rpm for 20 hours in Hitachi RPS 28-2 rotor. The thus precipitated RNA was collected and dissolved in 10 ml of a solution : 0.1% sodium laury/sulfate, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5). The RNA was extracted with phenol-chloroform and recovered by ethanol precipitation.

The thus obtained RNA (about 3.95 mg) was dissolved in 1 ml of a solution : 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. The solution was incubated at 65°C for 5 minutes, and 5 M NaCl (0.1 ml) was added. The resulting mbdure was subjected to chromatography on an oligo(dT) cellulose column (column volume of 0.5 ml, P-L Biochemical). The thus adsorbed poly(A)-containing mRNA was eluted with a solution : 10 mM Tris-HCl (pH 7.5), 1 mM EDTA. There was obtained about 100 μg of poly(A)-containing mRNA.

The thus obtained poly(A)\* mRNA (10  $\mu$ g) was dissolved in 50  $\mu$ l of RT buffer : 20 mM Tris-HCl (pH 8.8), 0.1 M KCl, 12 mM MgCl<sub>2</sub>, 2 mM MrCl<sub>2</sub>. To this solution, there was added 8  $\mu$ g of random primer d(N)<sub>6</sub> (P-L Biochemical). The resulting mixture was heated at 95°C for 3 minutes to denature the materials, which was then cooled gradually to room temperature to anneal the random primer with the mRNA. To the annealed mixture, there were aded 10 mM 4NTP (10  $\mu$ l) and reverse transcriptase (225 units) from TAKARA SHUZO (Japan), and then water was added so as to make the total volume of the mixture to 100  $\mu$ l. Reaction was allowed to proceed at 42°C for one hour.

To the reaction mixture (50  $\mu$ l), there were added 10 mM NAD (2  $\mu$ l), 10 mM 4dNTP (10  $\mu$ l), RNase H (5 units), <u>E. coli</u> ligase (1 unit), <u>E. coli</u> DNA polymerase I (6.3 units), and 10x T4 DNA ligase buffer (10  $\mu$ l; 0.1 M Tris-HCl, pH 7.5, 0.1 M DTT, 60 mM MgCl<sub>2</sub>) to make the total volume to 100  $\mu$ l. The mixture was allowed to react at 37°C for one hour to synthesize a double stranded DNA.

The thus obtained double stranded DNA was extracted with an equal volume of water-saturated phenol. Phenol in the aqueous layer was removed with the aid of ether followed by ethanol precipitation. The precipitate thus obtained was dissolved in 50  $\mu$ l of water, and 10x T4 DNA polymerase buffer (10  $\mu$ l; 0.33 M Tris-acetic acid, pH 7.9, 0.68 M potassium acetate, 0.1 M magnesium acetate, 5 mM DTT), 10 mM 4dNTP (10  $\mu$ l), and T4 DNA polymerase (6 units) were added to make the total volume to 100  $\mu$ l. The mixture was reacted at 37°C for one hour. There was obtained a double stranded DNA having blunt ends, which was then extracted with phenol to remove proteins and purified by ethanol precipitation as described above. The thus purified DNA was then air dried.

To the purified DNA, there were added 50 mM Tris-HCl (pH 7.5), 1 mM Na<sub>z</sub>EDTA, 5 mM DTT (20  $\mu$ ), 100  $\mu$ M S-adenosyi-L-methionine (2  $\mu$ I), and 1.8 mg/ml EcoRI methylase (0.2  $\mu$ I). Reaction was effected at 37°C for 15 minutes, whereby methylating the EcoRI restriction enzyme cleavage site on the DNA fragment. The reaction mixture was then heated at 70°C for 15 minutes to deactivate the enzyme.

To the reaction mixture, there was added 3'-phosphorylated EcoRI linker (GGAATTCC) in an amount of

100 molecules thereof per molecule of the synthetic DNA. There were further added 10x T4 DNA ligase buffer (5 μl; 0.5 M Trls-HCl, pH 7.5, 60 mM MgCl<sub>2</sub>, 10 mM DTT), 0.1 M ATP (5 μl), and T4 DNA ligase (5 units) to make the total volume to 50 μl. The resulting reaction mixture was reacted at 4°C for 16 hours followed by heating at 70°C for 10 minutes to deactivate the enzyme. Then, 10x EcoRi buffer (10 μl; 15 M Tris-HCl, pH 7.5, 0.5 M NaCl, 60 mM MgCl<sub>2</sub>), and EcoRi (100 units) were added to make the total volume to 100 μl, and the reaction mixture was reacted at 37°C for 2 hours to cut the linker. The reaction mixture was passed through Blo Gel A-50 (0.2 cm × 32 cm, Bio RAD). Eution was effected by a buffer: 10 mM Tris-HCl (pH, 7.5), 6 mM MgCl<sub>2</sub>. Excess EcoRi linker was removed and, thus, a double stranded cDNA having EcoRi sites at both ends thereof was purified.

To the thus obtained double stranded cDNA fragment having EcoRl sites at both ends, there were added gt 11 DNA (10  $\mu$ g) cleaved with EcoRl, 10x T4 DNA ligase buffer (10  $\mu$ l) as described above, 0.1 M ATP (10  $\mu$ l), and T4 DNA ligase (10 units) to make the total volume to 100  $\mu$ l. The mixture was reacted at 4°C for 16 hours. Thus, said double stranded cDNA fragment was inserted into  $\lambda$ gt 11 DNA.

The  $\lambda$  phage packaging kit (PROMEGA, Biotech) was used to introduce said DNA into  $\lambda$  phage particle. The procedures for packaging were effected according to the instructions of the kit.

The  $\lambda$ gt 11 phage having said DNA packaged thereinto was used to transfect <u>E. coli</u> strain Y1090 to form plaques according to the conventional Tomizawa et al. methods described in "Experimental Procedures for Bacterlophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). Among about 200,000 plaques, one positive clone was selected by immunological screening as described hereinbelow. A monoclonal antibody used in the immunological screening was prepared by the method described in Japanese Patent Application Laving-open (KOKAI) No. 91328/88.

E. coli Y1090 (R.A. Young et al., Pro. Natl. Acad. Sci. U.S.A., 80, 1194-1198 (1983), which had been transfected with Agt 11, was inoculated in a petri dish together with soft agar held at 42°C. The transfected cell was allowed to stand at 42°C for 5 hours. A nitrocallulose filter (S & S, BA-83, pore size of 0.2 µm) containing 10 mM iPTG was placed on the cell in the dish and incubation was effected at 37°C for 3-4 hours. This nitrocellulose filter was lightly rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed in the TBS buffer (400 mi) containing 3% gelatine and shaked at 40°C for one hour. Thus, the nitrocellulose filter was blocked. Then, a monoclonal antibody (OD<sub>280</sub> = 4.3) directed to a non-A non-B hepatitis-specific antigen was added to TBS buffer containing 1% getatine with a dilution of 1/400. This mbeture was put into a vinyl bag together with the filter in a proportion of 2 mi of the mbdure per filter, and reaction was allowed to proceed at room temperature for 16 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes. A labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Blo Rad) was added to TBS buffer containing 1% gelatine with a dilution of 1/1,000. This mixture and the filter were put into a vinyl bag with a proportion of 2 ml of the mixture per filter. Reaction was allowed to proceed at room temperature for 2 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes, in the same manner as described above. Color development was effected by dipping the filter and 4-chloro-1-naphthol (12 mg, Blo Rad) into 20 ml of TBS buffer containing hydrogen peroxide. After completion of the color development, the filter was thoroughly washed with water and put into a vinyl bag containing water. The bag was stored in a dark and cold place.

Thus, one positive plaque was obtained. The plaque was three times subjected to single plaque isolation. In each time, immunological screening was effected in the same manner as described above, verifying that the plaque was in fact positive.

40

The phage was then cultured in a large scale to purify the DNA in the following manner: First, <u>E. coli</u> Y1090 was cultured overnight in 10 ml of NZ medium prepared by adding NZ amine (10 g), NaCl (5 g) and 5 mM MgCl<sub>2</sub> to one liter of water followed by adjusting the pH to 7.2. The culture (1 ml) was transfected with the phage, with the m.o.i. (multiplicity of infection) being 0.1. The transfected culture was allowed to stand at 37°C for 10 minutes and then transferred to one liter of NZ medium. Shaking culture was effected at 37°C for 7-8 hours until the cells were lysed. Chloroform (5 ml) was added to the culture and shaking was continued for additional 30 minutes. The culture was subjected to centrifugation at 6,500 rpm for 10 minutes to remove cell debris.

NaCl (29 g) and polyethylene glycol (70 g) were added to and thoroughly dissolved in the obtained supernatant, and the solution was allowed to stand at 4°C overnight. The precipitate was collected by centrifugation at 6,500 rpm for 20 minutes, drained thoroughly, and dissolved in 20 ml of TM buffer: 10 mM Tris-HCl (pH 7.5), 5 mM MgCl<sub>2</sub>. DNase I and RNase A were added to the solution, both with a concentration of 10 μg/ml, and the reaction was effected at 37°C for one hour. Chloroform (20 ml) was then added to the reaction mbdure and stirred; thus, polyethylene glycol was distributed in the chloroform layer which was then separated from the aqueous layer. This aqueous layer was ultra-centrifuged at 28,000 rpm for 60 minutes. Thus, a pellet of phage particles was obtained.

This pellet was dissolved in TM buffer (1 ml) and subjected to CsCl density gradient centrifugation at 33,000

rpm for 20 hours. The resultant fraction containing the phage particles ( $\rho$  = 1.45-1.50) was dialyzed ovemight against TM buffer. Proteinase K was added to the dialyzate in an amount of 100  $\mu$ g/ml and reaction was effected at 37°C for one hour. Thereafter, an equal volume of water-saturated phenol was added and phenol-extraction was gently effected. After centrifugation at 6,500 rpm for 10 minutes, the aqueous layer was removed, put into a dialyzis tube, and dialyzed overnight against water at 4°C. Thus, about 5 mg of DNA was obtained.

Cleavage reaction of this DNA (100 μg) with EcoRl (100 units) in the aforementioned buffer (100 μl) at 37°C revealed that two cDNA segments of 390 bp and 345 bp were inserted into the phage DNA.

These two EcoRI fragments were re-cloned into EcoRI site of a cloning vector pUC 119. Base sequences of these DNA fragments were determined by the dideoxy method using commercially available primers CAG-GAAACAGCTATGAC and AGTCACGACGTTGTA, respectively. The base sequence of the linking portion between these two DNA fragments was similarly determined by cutting this cDNA fragment at BamHi and EcoRV sites present therein with corresponding specific restriction enzymes, inserting the resulting BamHi-EcoRV DNA fragment between BamHi and Smal sites of the plasmid pUC 119, and sequencing the fragment by the dideoxy method.

The base sequence of said cDNA fragment is shown in Fig. 3. This was a partial cDNA fragment of a gene coding for an antigenic protein specific to non-A non-B hepatitis.

### EXAMPLE 2: Preparation of cDNA Containing the Full Length Gene Sequence

Messenger RNA was prepared as described in Example 1 and cDNA was synthesized using Okayama vector according to the conventional method described in Molecular Cloning, p. 211 et seq. The procedures used to synthesize cDNA were as follows:

To 300 µl of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 10 mM NaCl), there were added 400 µg of pCDV 1 (Ckayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 500 units of Kpnl (TAKARA SHUZO, Japan), all restriction enzymes used hereinafter having been manufactured by TAKARA SHUZO (Japan) unless otherwise noted. Reaction was effected at 37°C for 6 hours to cut the plasmid at Kpnl site therein. After phenol-chloroform extraction, ethanol precipitation was effected to recover DNA.

The DNA (about 200 µg) cleaved with <u>KpnI</u> was added to 200 µl of a solution which was obtained by adding dTTP in a concentration of 0.25 mM to a buffer (TdT buffer): 40 mM sodium cacodylate, 30 mM Tris-HCl (pH 6.8), 1 mM CaCl<sub>2</sub>, 0.1 mM dithiothreitol (DTT). Further, 81 units of terminal deoxynucleotidyl transferase (TdT, manufactured by P-L Biochemicals) was also added. Reaction was effected at 37°C for 11 minutes. Thus, a poly(dT) chain (about 67 deoxythymidylic acid residues) was added to the 3' end at the <u>KpnI</u>-cleaved site of pCDV 1. After phenol-chloroform extraction and ethanol-precipitation, about 100 µg of pCDV 1 DNA to which poly(dT) chain had been added was recovered from the reaction mixture.

The thus obtained DNA was added to 150 μl of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 100 mM NaCl), and <u>Hpal</u> (360 units) was also added, followed by reaction at 37°C for 2 hours. The reaction mixture was subjected to electrophoresis on agarose gel to separate and recover about 3.1 Kbp DNA fragment. Thus, there was obtained about 60 μg of poly(dT)-containing pCDV 1.

The thus obtained DNA was dissolved in 500 μl of a solution (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), incubated at 65°C for 5 minutes, and cooled on ice. After adding 5 M NaCl (50 μl), the mixture was subjected to chromatography on oligo(dA) cellulose column (Colaborative Research). DNA having a poly(dT) chain of sufficient length was adsorbed on the column and eluted with a solution : 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Thus, there was obtained 27 μg of pCDV 1 to which poly(dT) chain had been added, abbreviated hereinafter as vector primer.

A linker DNA was prepared in the following manner: To 200 μ of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 50 mM NaCl), there were added about 14 μg of pL 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 50 units of Pstl. Reaction was effected at 37°C for 4 hours to cut the pL 1 DNA at Pstl sits. Phenol-chloroform extraction and ethanol precipitation of the reaction product gave about 13 μg of pL 1 DNA cleaved at Pstl sits.

The thus obtained DNA (about 13 µg) was added to 50 µl of the TdT buffer containing dGTP at a final concentration of 0.25 mM, and 54 units of TdT (P-L Blochemicals) was also added. The mixture was incubated at 37°C for 13 minutes to add a (dG) chain (about 14 deoxyguanytic acid residues) to the 3' end at the Psti-cleaved site of pL 1. After phenol-chloroform extraction, DNA was recovered by ethanol precipitation.

The thus obtained DNA was added to 100 µl of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 80 mM NaCl), and 80 units of <u>Hindill</u> was also added. The mixture was incubated at 37°C for 3 hours to cut the pL 1 DNA at <u>Hindill</u> site. The reaction product was fractionated by agarose gel electrophoresis. About 0.5 Kb DNA fragment was recovered by the DEAE paper method: Dretzen et al., Anal. Biochem., <u>112</u>, 295 (1981). Thus, there was obtained an oligo(dG) chain-containing linker DNA, hereinafter abbreviated simply as linker DNA.

The aforementioned poly(A)\* RNA (about 2 μg) prepared in the same manner as in Example 1 and the vector primer (about 1.4 μg) were dissolved in 22.3 μl of a solution : 50 mM Tris-HCl (pH 8.3), 8 mM MgCl<sub>2</sub>, 30 mM KCl, 0.3 mM DTT, 2 mM dNTP (dATP, dTTP, dGTP and dCTP) and 10 units of ribonuclease inhibitor (P-L Blochemicals). To the solution, there was added 10 units of reverse transcriptase manufactured by SEIKAGAKU KOGYO (Japan). Incubation was effected at 37°C for 40 minutes to synthesize a DNA complementary to the mRNA. After phenol-chloroform extraction and ethanol precipitation, the vector primer DNA to which a double stranded RNA-DNA had been added was recovered.

The thus obtained vector primer DNA containing RNA-DNA double stranded chain was dissolved in 20 μl of TdT buffer containing 60 μM dCTP and 0.2 μg poly(A). After adding 14 units of TdT (P-L Biochemical), the mbxture was incubated at 37°C for 8 hours to add a (dC) chain of 12 deoxycytidylic acid residues to the 3′ end of the cDNA. The reaction product was extracted with phenol-chloroform and precipitated with ethanol to recover a cDNA-vector primer DNA to which a (dC) chain had been added.

The thus obtained (dC) chain-containing cDNA-vector primer DNA was dissolved in 400 µl of a solution (10 mM Trie-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 60 mM NaCl), and 20 units of <u>Hindlil</u> was also added. The mixture was incubated at 37°C for 2 hours to cut the DNA at <u>Hindlil</u> site. The reaction product was extracted with phenol-chloroform and precipitated with ethanol. Thus, there was obtained 0.5 pmole of a (dC) chain-containing cDNA-vector primer DNA.

The thus obtained (dC) chain-containing cDNA-vector primer DNA (0.08 pmole) and the aforementioned linker DNA (0.16 pmole) were dissolved in 40 μl of a solution: 10 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 1 mM EDTA. The resulting solution was incubated at 65°C for 10 minutes, at 42°C for 25 minutes, and then at 0°C for 30 minutes. The reaction mixture was adjusted to 20 mM Tris-HCl (pH 7.5), 4 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M KCl and 0.1 mM β-NAD in a total volume of 400 μl.

To the reaction mixture, there was added 10 units of E. coli DNA ligase (New England Biolabs), followed by incubation overnight at 11°C. After adjusting the concentrations of dNTP and β-NAD in the reaction mixture to 40 μM and 0.15 mM, respectively, by supplementing necessary reagents, 5 units of E. coli DNA ligase, 7 units of E. coli DNA polymerase I (P-L Biochemicals) and 2 units of E. coli ribonuclease H (P-L Biochemicals) were added to the reaction mixture. The mixture was incubated at 12°C for one hour and then at 25°C for one hour.

In the course of the above reactions, a recombinant DNA containing the cDNA was cyclized and the RNA portion of the RNA-DNA double stranded chain was substituted by DNA. Thus, a desired recombinant plasmid containing a complete double-stranded DNA was produced.

The recombinant plasmid was used to transform competent cells of <u>E. coli</u> strain MC1064 prepared by conventional methods. Approximately 50,000 transformants were fixed on a nitrocellulose filter. These colonies were screened according to the colony hybridization method described in Molecular Cloning, Cold Spring Harbor Laboratory, p. 329 et seq. (1982) using the cDNA fragment obtained in Example 1 as a <sup>32</sup>P-labelled probe. Thus, three clones showed strong hybridization at 42°C.

These positive clones were analyzed in detail by Southern method: J. Mol. Biol., <u>98</u>, 503 (1975). There was obtained the desired full length cDNA of a gene coding for an antigenic protein specific to non-A non-B hepatitis. The base sequence of the cDNA is shown in Fig. 4.

The expression vector containing the full length cDNA was designated as pCDVCL-I.

### EXAMPLE 3: Preparation of Expression Vector and Transformant and Expression of Specific Antigen

- A. Preparation of Expression Vector and Transformant
- I) Modification of N-terminus (Fig. 5):

i) in 100 μl of a buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl<sub>2</sub>), pCDVCL-I (5 μg) was digested with PvuI (10 units) at 37°C for 2 hours. The reaction mixture was heated at 75°C for 15 minutes to deactivate the enzyme, distyzed against water, and dried. The cleaved plasmid DNA was treated with T4 DNA polymerase (4 units) in 40 μl of a system: 33 mM Tris-acetic acid (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate and 0.5 mM dithiothreitol, to which 2 mM 4-deoxytriphosphate had been added; thus, the 3′ protruding end of the plasmid DNA was filled in to produce a blunt end. The thus treated mixture was heated at 70°C for 10 minutes to deactivate the enzyme, distyzed against water, and dried. The thus obtained plasmid DNA was then stored in the form of an aqueous solution (50 μl). This plasmid DNA fragment is hereinafter designated as Fragment i.

ii) On the other hand, pCDVCL-I (20 μg) was digested with Ncol and Hindill (each 20 units) at 37°C for 2 hours in 100 μl of a buffer : 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 6 mM MgCl<sub>2</sub>. The plasmid DNA was

subjected to 5% acrylamide gel electrophoresis at 10 V/cm for 1.5 hours in a buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA. The gel was stained with 0.05% aqueous ethidium bromide solution and two gel silices corresponding to DNA fragments of larger molecular weights were excised from the gel under ultraviolet radiation at 340 nm. The gel slices were crushed by means of a glass rod, suspended into 4 mi of a buffer for DNA extraction (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% sodium laury/sulfate), and allowed to stand overnight at 37°C to extract DNA from the gel. The materials were subjected to centrifugation at 10,000 rpm for 15 minutes to eliminate larger gel pieces, and passed through a glass filter to remove smaller gel pieces. The DNA was purified by effecting ethanol precipitation three times and stored in the form of an aqueous solution (200 µl). This plasmid DNA fragment is hereinafter designated as Fragment II.

iii) A primer of the DNA portion to be modified as shown below (51 bases) was synthesized by a DNA synthesizer, NIKKAKI (Japan), Applied Biosystem MODEL 380A. The synthesized DNA was overnight reacted with concentrated aqueous ammonia at 55°C to deprotect and purified by reversed HPLC before use.

HindIII

Primer ACAACAGATCTAAGCTTATGGCAGTTACAACAAGATTAA

(Original sequence) (-----A-A-G------G----TC-C--G
CATGGTTGCATG wherein x represents a base substitution.

The synthetic primer (150 pmole) was treated with T4 polynucleotide kinase (20 units) in 10  $\mu$ l of a kinase buffer (50 mM Tris-HCi, pH 8.0, 10 mM MgCl $_2$  5 mM dithiothreitol) to phosphorylate the 5' end thereof. Iv) Fragment I (0.05 pmole), Fragment II (0.05 pmole) and 5'-phosphorylated primer (45 pmole) were added to 12  $\mu$ l of 5x polymerase-ligase buffer (0.5 M NaCl, 32.5 mM Tris-HCl, pH 7.5, 40 mM MgCl $_2$  5 mM  $\beta$ -mercaptoethanol) to make the total volume of the mixture 34.8  $\mu$ l. The mixture was boiled at 100°C for 3 minutes, immediately after which it was placed in a thermostat at 30°C and allowed to stand for 30 minutes. The mixture was allowed to stand at 4°C for 30 minutes and then on ice for 10 minutes to form a heterodup-lex.

To an aqueous solution (11.6  $\mu$ ) containing the heteroduplex, there were added 2.5 mM 4-deoxynucleotide triphosphate (2  $\mu$ I), 10 mM ATP (2  $\mu$ I), Klenow enzyme (2 units) and T4 DNA ligase (0.5 units) to form a mixture of 20  $\mu$ I in total volume. The mixture was reacted overnight at 16°C to cyclize the DNA.

An aqueous solution (2 µl) containing the circular DNA was used to transform E. coli HB101 strain according to conventional methods. Plasmids were separated from the transformant and purified in conventional manners. The plasmid was cleaved with restriction enzyme Hindlil and subjected to 5% acrylamide gel electrophoresis. Thus, two separate fragments were collected as desired modified, variant plasmids. Since resulting variant plasmids might often be admixed with original wild-type plasmids, the thus obtained variant plasmids were again employed to transform E. coli HB101 so as to purify the plasmid.

Thus, a purified plasmid pCV44H was obtained (Fig. 5).

- (5 II) Modification of C-terminus (Fig. 6):
  - i) Plasmid pCDVCL-I (5  $\mu g$ ) was treated in the same manner as in I) i) described above to produce Fragment I.
  - ii) Plasmid pCDVCL-I (20  $\mu$ g) was treated in the same manner as in I) II) described above except that Ncol and NsII (each 5 units) were employed. Thus, Fragment II was produced.
  - iii) In the same manner as in I) iii) described above, the following primer (46 bases) was synthesized and the 5' end thereof was phosphorylated.

55

50

10

15

25

	Primer		GCACAAGG	II <u>lpB</u> <u>'Otkākā</u> TKAAAAA XXXXX	Sali TGTCGACGGTT XXXXXXX	CACGTA
5	(Original se	equence)	(	AGATAT	GTGAA*A	
			AATTTCC	wherein x rep substitution	resents a band * repre	ase Sents
10			)	an addition.		

iv) The Fragment I and II and the 5'-phosphorylated primer obtained above in II) i) to iii) were treated in the same manner as in I) iv) described above. Thus, plasmid pCV44B was obtained (Fig. 6).

III) Introduction of cDNA coding for specific antigen into expression vector (Fig. 7):

i) in 100 µl of a buffer H (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl<sub>2</sub>), 10 µg (about 3 pmole) of pCV44H was cut with Hindill (20 units) and Sacl (20 units) at 37°C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis. Thus, a 467 bp DNA fragment coding for the N-terminus of the specific antigen was separated and purified. This fragment is hereinafter designated as Fragment N

ii) In 100 µl of the buffer H, 10 µg (about 3 pmole) of pCV44B was cleaved with <u>Bglll</u> (20 units) and <u>Sacl</u> (20 units) at 37°C for 2 hours. The reaction mbdure was subjected to 5% acrylamide gel electrophoresis to isolate and purify a 836 bp DNA fragment coding for the C-terminus of the specific antigen. The thus obtained fragment is hereinafter designated as Fragment C.

iii) In 20 μl of buffer H, 2 μg (about 1 pmole) of an expression vector pUSAH was cut with <u>Hindlil</u> (2 units) and <u>Bgill</u> (2 units) at 37°C for 2 hours. The reaction mixture was extracted with an equal volume of water-saturated phenol to remove proteins. After extracting the phenol with ether, the reaction mixture was dialyzed against water to desait, and concentrated by a vacuum pump. Thus, there was obtained 10 μl of an aqueous solution containing an expression vector fragment HB.

iv) Fragment N (0.5 pmole), Fragment C (0.5 pmole) and the expression vector fragment HB (0.1 pmole) were mixed and reacted with T4 DNA ligase (1 unit) at 4°C for 16 hours in 10 μi of a buffer (10 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, 6 mM MgCl<sub>2</sub>, 1 mM ATP). The reaction mixture (3 μl) was used to transform commercially available <u>E. coli</u> JM109 competent cell according to conventional methods. The resulting transformants were selected in L broth plate (bactopeptone 10 g, yeast existract 5 g, NaCl 10 g, agar 15 g per liter) containing 20 μg/ml ampicillin. Thus, there was obtained an expression vector pCZ44 containing the specific antigen gene inserted thereinto (Fig. 7).

### B. Expression of Specific Antigen

20

25

30

35

40

E. coli strain JM109 possessing pCZ44 was cultured overnight at 30°C in L broth. The culture was inoculated in a fresh L broth with a dilution of 1/50 and cultured with shaking at 30°C for 2 hours. After IPTG (isopropylithio-β-D-galactopyranoside) was added to the medium in a concentration of 2 mM, shaking culture was continued at 30°C for further 3 hours. The cells were collected by centrifugation at 6,500 rpm for 10 minutes and suspended in a buffer (0.9% NaCl, 10 mM Tris-HCl, pH 7.5) to store.

#### C. Verification of Expression of Specific Antigen

The thus obtained cell culture (0.3 ml) was subjected to 10% SDS polyacrylamide gel electrophoresis at 120 V for one hour in a buffer (Tris 3g/l, glycine 14.4 g/l, 0.1% SDS). The gel was removed, placed on a nitrocellulose filter, interposed between filter papers and electrophoresed at 5 V/cm, 4°C in a buffer (Tris 3g/l, glycine 14.4 g/l) to transferred proteins in the gel onto the nitrocellulose filter. The nitrocellulose filter was rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed into 400 ml of TBS buffer containing 3% gelatine and shaked at 40°C for one hour to block the nitrocellulose filter.

To TBS buffer containing 1% gelatine, there was added a monocional antibody directed to a non-A non-B hepatitis-specific antigen ( $OD_{280} = 4.3$ ) with a dilution of 1/400. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 16 hours. The reaction mixture was washed three times with 400 ml of TBS buffer con-

taining 0.05% Tween 20 for 10 minutes. -

To TBS buffer containing 1% gelatine, there was added a labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad), with a dilution of 1/1000. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 2 hours. The reaction mixture was washed three times with 400 ml of TBS buffer containing 0.05% Tween 20 for 10 minutes.

Color formation was effected by immersing the filter into 20 ml of TBS buffer containing 12 mg of 4-chloro-1-naphthol (Bio Rad) and hydrogen peroxide. After completion of color formation, the filter was thoroughly washed with water, put into a vinyl bag containing water, and stored in a dark and cold place.

Such a test effected showed that a protein reacting with the monoclonal antibody was found at the same position (44 Kd) as found in the case of the specific antigen derived from infected chimpanzee liver. This verifies that such a specific antigen can be in fact expressed in E coli. The invention thus also relates to a process for the in vitro diagnosis of NON-A NON-B hepatitis, which comprises contacting a liver sample and/or a serum sample taken from a patient possibly infected with a NON-A NON-B hepatitis with the protein whose formula appears in claim 3 hereafter or a part thereof for a time and under conditions sufficient to allow for the production of a complex between said protein or part thereof with the antibodies contained in the patient sample and detecting the presence of the immunologic complex, particularly when the patient is suffering from NON-A NON-B hepatitis.

Any part of said protein, or any recombinant, produced by genetic engineering and including the aminoacid sequence of said protein or part of said protein can be substituted for above-said protein, it being understood that the said recombinant protein or part of said protein are specifically recognized by the same antibodies as those which recognize said protein.

In other words the invention relates to all recombinant proteins or protein fragments which bind to antibodies contained in a liver extract or serum sample, or both, and originating from a patient suffering from NON-A NON-B hepatitis.

The invention also relates to a process for detecting in vitro an infection by a NON-A NON-B hepatitis virus, which process comprises contacting the DNA of claim 4, or a fragment thereof, under suitable hybridization conditions, with a sample of liver extract and/or serum sample originating from the patient to be diagnosed and in which the nucleic acid components had previously been made accessible to hybridization, to form a hybridization product between said DNA of claim 4 (probe) and the viral DNA of a NON-A NON-B hepatitis B virus, and detecting said hybridization product, particularly in the case where the patient is indeed infected with a NON-A NON-B virus.

#### 35 Claims

50

1. A DNA fragment which contains a base sequence coding for an antigenic protein specifically occurring in a host affected with non-A non-B hepatitis, said protein comprising the whole or a part of the amino acid sequence represented by the formula:

5	Met	Ala	Val	Thr	Thr	Arg	Leu	Thr	Trp	10 Leu	His	Glu	Lys	Ile	Leu
	Gln	Asn	His	Phe	20 Gly	Gly	Lys	Arg	Leu	Ser	Leu	Leu	Tyr	Lys	30 Gly
10			His							40					
			Gly		50										60
15			Ala							70					
			Ile		80										00
20										100					
			Gly		110									_	
25	مند	778	Tyr	ABII	Ser	PED	THE	ABI	Phe	130	Ile	Asp	Gly	yrg	Asn
30	Arg	Lys	Val	Ile	Met 140	Asp	Leu	Lys	Thr	Met	Glu	λsπ	Leu	Gly	
<b></b>	Ala	Gln	λsn	Суз	The	Ile	Ser	Ile	GIn		Tyr	Glu	Val	Phe	150 Arg
35	Сув	Glu	Asp	Ser		Asp	Ğlu	Arg	Lys	160 Ile	Lys	Gly	Val	Ile	Glu
	Leu	Arg	Lys	Ser	170 Leu	Leu	Ser	Ala	Leu	Arg	The	TYE	Glu	Pro	180 Tyr
40	Gly	Ser	Leu	Val	Gln	Gln	Ile	ÀΣα	Ila	190		T.=	<i>(</i> 23 ••	- B	77.

=

200 210 Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu Ser Glu Lys Glu Gly Gly Leu Cys Met Asp Asp Ile Ser Tyr Ile Leu Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu Ser Ile Lys Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu Lys Asp Arg Ile His Cys Val Ala Phe Val Phe Asp Ala Ser Ser Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile Arg Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu Thr His Val. Asp Ser Met Asp Leu Ile Thr Lys Gly Asp Leu Ile 35 Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys Leu Glu Glu Val Gin Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu 45 Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala Ala Asp Asp Phe Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu Arg Glu Glu 50 Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*.

<sup>2.</sup> The DNA fragment in accordance with claim 1, in which the base sequence comprises the whole or a part of the base sequence represented by the formula:

				1	.0			20			30			4	0
		ATG									TIG	CAT	GAA	AAG	ATC
5	; 3 '														
				50			60			7	'n			80	
		CTG				TTT									TAT
10									_						
10						10									
		AAG	GGT	AGT	GTC	CAT	GGA	TTC	CAT	AAT	GGA	GTT	TTG	CTT	GAC
		1:	30		1	140		•	150			10	50		
15		AGA	TGT	TGT	AAT	CAA	GGG	CCT	ACT	CTA	ACA	GTG	ATT	TAT	AGT
		170			180			19	90		:	200			210
		170 GAA	GAT	CAT	180 ATT	ATT	GGA	19 GCA	O TAT	GCA	GAA	200 GAG	GGT	TAC	210 CAG
20		170 GAA	GAT	CAT	180 ATT	ATT	GGA	GCA	O TAT	GCA	GAA	200 GAG	GGT	TAC	210 CAG
20		170 GAA	GAT	CAT	ATT	ATT	GGA	GCA	TAT	GCA	GAA	GAG	GGT 	TAC	CAG
20		GAA	GAT	CAT	ATT 	ATT	GGA	GCA 	TAT	GCA	GAA  240	GAG	GGT	TAC	CAG 
20		GAA GAA	GAT 	CAT  2: AAG	ATT 20 TAT	ATT	GGA	GCA  230 ATC	TAT	GCA	GAA  240	GAG	GGT	TAC	CAG 
-		GAA GAA	GAT 	CAT  2: AAG	ATT 20 TAT	ATT	GGA	GCA  230 ATC	TAT	GCA	GAA  240	GAG	GGT	TAC	CAG 
20 25		GAA GAA	GAT AGA	CAT  2: AAG	ATT  20 TAT	GCT	TCC	GCA 230 ATC	ATC	CTT	240 TIT	GAG	CTT	Z: CAA	CAG 
-		GAA	AGA	CAT  Z  AAG  260	ATT 20 TAT	GCT	TCC	GCA  230 ATC	ATC	CTT2	GAA 240 TIT	GAG	GGT  CTT	TAC  2: CAA 	CAG 50 GAG
-		GAA	AGA	CAT  Z  AAG  260	ATT 20 TAT	GCT	TCC	GCA  230 ATC	ATC	CTT2	GAA 240 TIT	GAG	GGT  CTT	TAC  2: CAA 	CAG 
25		GAA GAA ACT	AGA	ZAT AAG AATT	ATT 20 TAT TCA	GCT GAA	TCC 270	GCA 230 ATC	ATC	CTT 2 GGA	GAA 240 TIT B0 CTA	GCA	CTT ACA	22 CAA  290 CCA	CAG  50 GAG
-		GAA GAA ACT	AGA AAA 300	Z: AAG  260 ATT	ATT 20 TAT TCA	GCT GAA	GGA TCC 270 TGG	GCA  230 ATC 	ATC CTA	GCA CTT GGA 320	GAA 240 TIT  80 CTA	GCA	CTT ACA 330	290 CCA	CAG GAG GAA
25		GAA GAA ACT	AGA AAA 300	Z: AAG  260 ATT	ATT 20 TAT TCA	GCT GAA	GGA TCC 270 TGG	GCA  230 ATC 	ATC CTA	GCA CTT GGA 320	GAA 240 TIT  80 CTA	GCA	CTT ACA 330	290 CCA	CAG 50 GAG

5	AAT	TTC	CAG	ATA	GAT	GGA	AGA	360 AAT	AGA	AAA 	37 GTG	O ATT	ATG	GAC
	380 TTA	AAG	ACA	ATG	GAA	AAT	CTT	GGA	CTT	GCT	CAA		TGT	420 ACT
10	ATC	TCT	ATT	CAG	GAT	TAT	GAA	GTT	TTT	CGA	TGC	GAA	GAT	TCA
15,	CTG	GAC	70 GAA	AGA	AAG	ATA	AAA	GGG	49 GTC	O ATT	GAG	CTC	AGG	AAG
20		TTA	CTG	TCT	GCC	TTG	AGA	ACT	TAT	GAA	CCA	TAT		
25	CTG	GTT	CAA	CAA	ATA	CGA	ATT	CTG	CTG	CTG	GGT	CCY		GGA
		GGG	AAG	TCT	AGC	TTT	TTC	AAC	TCA	GTG	AGG		GTT	
30	CAA		CAT	GTA	ACG	CAT	CAG	GCT	TTG	GTG	GGC		AAT	ACA
35	ACT	GGG	680 ATA	TCT	GAG	690 AAG	TAT	AGG	ACA	TAC	TCT	ATT	710 AGA	GAC
40		AAA	GAT	GGC	AAA	TAC	CTG	CCA	TTT	ATT	CTG	750 TGT	GAC	TCA
45	CIG	GGG	CTG	AGT	GAG	AAA	GAA	GGC	GGC	CTG	TGC	ATG		GAC
50	800 ATA		TAC	810 ATC		AAC		20 AAC 			830 GAT			840 CAG
~	TTT	AAT									CAT	CAT	GAC	80 TAC

	አጥጥ ሮኔጥ ጥሮሮ ሮር	'A TCG CTG AAG	910 GAC AGA ATT CA	T TGT GTG GCA
5	TTT GTA TTT G	AT GCC AGC TCT	950 ATT GAA TAC TT	C TCC TCT CAG
10	970 ATG ATA GTA A	ag atc aaa aga	990 1 ATT CGA AGG GA	G TTG GTA AAC
15	1010 10 GCT GGT GTG G	TA CAT GTG GCT	30 1040 TTG CTC ACT CA	IT GTG GAT AGC
20	ATC GAT CTG A	TT ACA AAA GGT	1080 GAC CTT ATA GA	la ata gag aga
	TGT GTG CCT G	TG AGG TCC AAG	1120 CTA GAG GAA G	1130 CC CAA AGA AAA
25	CTT GGA TTT G	ict ctt tct gag	1160 ATC TCG GTG G	IT AGC AAT TAT
30	TCC TCT GAG 7	igg gag ctg gag	1200 CCT GTA AAG G	AT GTT CTA ATT
35	1220 1: CTT TCT GCT (	CTG AGA CGA AT	240 125 G CTA TGG GCT G	CA GAT GAC TTC
40	TTA GAG GAT	TTG CCT TTT GA	1290 G CAA ATA GGG A	1300 NAT CTA AGG GAG
	GAA ATT ATC	1320 AAC TGT GCA CA	1330 A GGA AAA AAA 3	3° 5°

wherein the sign "-" represents a base complementary to the base shown just above each sign.

- 5. The expression vector in accordance with claim 3, in which the promoter operates in a microorganism.
- 6. The expression vector in accordance with claim 3, in which the promoter operates in an eukaryote.
- 7. A transformant obtained by transforming a host with an expression vector in which a DNA fragment containing a base sequence according to claim 1 or 2 and coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis is introduced into a cloning site present downstream from a promoter of said vector.

<sup>3.</sup> An expression vector in which a DNA fragment containing a base sequence according to claim 1 or 2 and coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis is introduced into a cloning site present downstream from a promoter of said vector.

<sup>4.</sup> The expression vector in accordance with claim 3, in which the promoter is controllable by a regulatory factor.

8. The transformant in accordance to claim 7, in which the host is Escherichia coil or Bacillus subtills.

9. A process for producing an antigen occurring specifically in a host affected with non-A non-B hepatitis, comprising introducing a DNA fragment containing a base sequence according to claim 1 or 2 and coding for said specifically occuring antigen into a cloning site present downstream from a promoter of a vector for expression, introducing the expression vector contains said DNA fragment into a host, culturing said transformed, and collecting the produced and accumulated antigen.

10. Process for the In vitro diagnosis of non-A non-B hepatitis which comprises contacting a liver sample and/or a serum sample taken from a patient suspected of being infected with a non-A non-B hepatitis, with the protein whose sequence appears in claim 1, or a part thereof, for a time and under conditions sufficient to allow for the production of a complex between said protein or a part thereof with the antibodies contained in the patient sample and detecting the presence of the immunological complex, particularly when the patient is suffering from non-A non-B hepatitis.

### Revendications

20

25

30

1. Fragment d'ADN qui contient une séquence de bases codant pour une protéine antigénique rencontrée spécifiquement chez un hôte atteint de l'hépatite non-A non-B, ladite protéine comprenant la totalité ou une partie de la séquence d'acides aminés représentée par la formule :

											10					
	Het	Ala	V	/al	Thr	Thr	٨rg	Lau	Thr	Trp		His	Glu	Lys	Ila	Leu
5						20										30
	Gln	λs	1 1	iis	Phe	Gly	Gly	Lys	Arg	Lou	Ser	Leu	Leu	Tyr	Lys	Gly
											40			•		
10	Ser	Val	LI	Eis	Gly	Phe	His	λsn	Gly	Val		Leu	λsp	λrg	Cys	Cys
						50										60
	λsn	Gl	3 (	Gly	Pro		Leu	Thr	Val	Ile	Ťyz	Ser	Glu	Asp	His	Ile
											70					
15	Ile	GI	y .	ملة	Tyr	ملة	Glu	Glu	Gly	Tyr		Glu	Arg	Lys	TyE	Ala
			•			80				_				_		90
	Ser	IL	•	Ile	Lou		Ala	Leu	Gla	Glu	Thr	Lys	Ile	Ser	Glu	IIP
20											100					
	Lvs	Le	u	Gly	Leu	TYE	The	Pro	Glu	Thr	100 Lou		Cys	Сув	Asp	Val
	-, -		_	•		_							-	•	•	
25	Ma	ī.v	<b>T</b>	Tvr	Ann	110 Ser		The	Asz	Phe	Gla	Ile	l lat	Gly	. Arc	120 181
		-,	•	-,-												
							_	_	•	_	130	-	_	_		_
	yri	L	7	Val	Ile	e Met	: As;	Let	r TA	i in	c Mei	: GII	1 ABI	Let	i eri	Leu
10						140										150
	<b>A</b>	s Gi	מו	λει	2 CA1	Thi	r Ile	e Sei	r Il	e Gl	r ye	i Kr	r Gl	u Val	L Pho	e Arg
						•					16	0				
15	Cyr	5 G.	Lu	-221	y Se:	r Lo	ı Ası	p Gl	i yr	g Ly	s II	e Ly	# G1	y Va	1 11	e Glu
						170	0									180
	Le	u A	Eg	Ly	s Se	r Le	u Le	u Sa:	r XI	a Lo	u Ar	gTh	r Ty	T Gl	u Pr	o Tyr
										٠	19	0				
v	Gl	y S	e I	Lo	u Va	1 G1	n Gl	n Il	o Yi	g Il	a La	u Le	u Le	u G1	y Pr	o Ila

0

5	(Gly	Ala	Gly	Lys	200 Ser	Ser	Phe	Phe	Asn	Ser	Val	Arg	Ser	Val	210 Phe
	Gln	Gly	His	Val	Thr	His	Gln	λla	Leu	220 Val	Gly	Thr	λsn	Thr	Thr
10	Gly	Ile	Ser	Glu	230 Lys	Tyr	λrg	The	Tyr	Ser	Ile	λrg	λsp	Gly	240 Lys
	λsp	Gly	Lys	Tyr	Leu	Pro	Phe	Ile	Leu	250 Cys	λзр	Ser	Leu	Gly	Leu
<b>15</b>	Ser	Glu	Lys	Glu	260 Gly	Gly	Leu	Суз	Ket	λsp	λзр	Ile	Ser	Tyr	270 Ile
20	Leu	λsn	Gly	λsn	Ile	λrg	λsp	λrg	Tyr	280 Gln	Phe	λsn	Pro	Ket	Glu
	Ser	Ile	Lys	Leu	290 Asn	His	His	λsp	Tyr	Ile	λsp	Ser	Pro	Ser	300 Leu
25				Ile						310					
				Phe	320										330
30				Leu						340					•
35				Asp	350										360
				Arg						370	)		_		
40				Leu	380	I									390
										400					
45				Ser	410										
_				Ala											
50	Leu	Glu	Asp	Leu		Phe	Glu	Gln	Ile	430 Gly	Asa	Lou	Arg	Glu	Glu
55	Ile	Ile	λsn	Cys	440 Ala	Gln	Gly	Lys	Lys	***.	•			-	_

<sup>2.</sup> Fragment d'ADN selon la revendication 1, dans lequel la séquence de bases comprend la totalité ou une partie de la séquence de bases représentée par la formule :

				1	10			20			30			4	0
	5'	ATG	GCA	GTG	λCλ	<b>ACT</b>	CGT	TTG	<b>ACA</b>	TGG	TTG	CAT	GAA	AAG	ATC
	3'														
5				50			60				70			20	
		CTG	CAA	AAT	CAT	TTT	GGA	GGG	MG	cee	/U (****	AGC	CTT	CTC	TAT
10			90			10	00		1	L10			120		
		AAG	GGT	AGT	GTC	CAT	GGA	TTC	CXI	AAT	GGA	GTT	ŢŢĞ	CTT	GAC
		1	30			140			150			10	60		
15		λGλ	TGT	TGT	AAT	CAA	GGG	CCT	ACT	CTA	ACA	GTG	ATT	TAT	AGT
		170			180			1	90			200			210
						ATT									
20															
								_							
				2	20			230			240			2	50
		Gλλ	λGλ	λλG	TAT	GCT	TCC	ATC	ATC	CII	TII	GCA	CII	CYY	GAG-
25														•••	
				260			270			2	80			290	
		ACT	AAA	ATT	TCA	GAA	TGG	AAA	CTA	GGA	CTA	TAT	ACA	CCA	GAA
30			200			•	10			330					•
		101	300			3	TU	-	-	<b>JZ</b> 0			330		
		ACA	CIG	TIT	TUT	TGT	JNU 	GIT	~~~	****	TAT	AAC	700	CCY	ACT

380 390 400 410 410 ATT TOT ACT ACT ACT ACT ACT ACT ACT ACT ACT AC		340	3:	50	360	370
380 390 400 410 420  TTA AAG ACA ATG GAA AAT CTT GGA CTT GCT CAA AAT TGT ACT  TA ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  450 460 460  CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  CTG GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC		AAT TTC	CAG ATA	gat GGA AGA	AAT AGA AAA G	TG ATT ATG GAC
10  ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  10  ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  11  12  ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  15  CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  20  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25  590  GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  30  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  31  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
TTA AAG ACA ATG GAA AAT CTT GGA CTT GCT CAA AAT TGT ACT  10  ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  15  CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  20  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25  S90  GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  30  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  31  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	5	200	300	4	00 41	n 420
430 440 450 450 460  ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  470 480 490 500  CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  CTG GGT AAG AGA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  CAA GGG CAT GTA ACG CAT CAA GCT TAT GAG TCT GTT TTC  ACT GGG AAA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  470 CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  20 AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  590 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  26 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  570 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  571 ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  470 CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  20 AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  590 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  26 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  570 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  571 ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC					,	
THE CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  20 AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550 560 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  CTG GTT CAA CAA ATA CGA ATT CTG CTG GTG CCA ATT GGA  GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  680 690 700 710  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	10					
CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  S50 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
20 CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  20 AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550 560 560 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25 S90 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  26 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  27 ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC		_				
20 AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550 560 570 570 580 580 570 580 580 580 580 580 580 580 580 580 58			470	480	490	500
AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  S50 CTG GTT CAA CAA ATA CGA ATT CTG CTG GGT CCA ATT GGA  S90 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	10					
25  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25  590  GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  30  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
25  AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  550  CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  25  590  GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  30  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC		510		520	530	540
25  CTG GTT CAA CAA ATA CGA ATT CTG CTG GGT CCA ATT GGA  25  590 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  30  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	20	AGC TTA				
CTG GTT CAA CAA ATA CGA ATT CTG CTG GGT CCA ATT GGA  590 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  CAA GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
CTG GTT CAA CAA ATA CGA ATT CTG CTG GGT CCA ATT GGA  590 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  CAA GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC		550	5	50	570	500
590 GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  CAC GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  640 650 660 670  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  35 680 690 700 710  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	25					
GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  640 650 660 670  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  35 680 690 700 710  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC				_		
CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  CAG GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC		590 66 <b>8</b> 666	600		10 6	630
640 650 660 670  CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  35 680 690 700 710  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  680 690 700 710  ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	30					
680 690 700 710 ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC			640	650	660	670
680 690 700 710 ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC						
ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC		440 000				
ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC	35					
		ACT GGG	ATA TCT	GAG AAG TAT	' AGG ACA TAC	TCT ATT AGA GAC
				***		
<b>720 730 740 750</b>		720	3	730	740	750
GGG AAA GAT GGC AAA TAC CTG CCA TTT ATT CTG TGT GAC TCA	40	GGG AN	A GAT GGC	AAA TAC CTO	CCA TTT ATT	CTG TGT GAC TCA
740 770 790 700		740	•	170	790	700
760 770 780 790 « CTG GGG CTG AGT GAG AAA GAA GGC GGC CTG TGC ATG GAT GAC	4E					
and dee one one one one one one one one one o	~~					
				_		
800 810 820 830 840						
ATA TCC TAC ATC TTG AAC GGT AAC ATT CGT GAT AGA TAC CAG	50					
850 860 870 880						
TTT AAT CCC ATG GAA TCA ATC AAA TTA AAT CAT CAT GAC TAC						
55	55		_			

	8	90	900		910	9:	20
	ATT GAT		TCG CTG		AGA ATT		GTG GCA
5							
	930 TTT GTA	TTT GAT	GCC AGC	TCT ATT	50 GAA TAC	960 TTC TCC 1	TCT CAG
10	970	9	80	990		1000	
	ATG ATA	GTA AAG	ATC AAA	AGA ATT	CGA AGG	GAG TTG	GTA AAC
40	1010	1020		1030	10	40	· 1050
15	GCT GGT				CTC ACT		GAT AGC
	<b>አ</b> ሞር <b>ፍ</b> ልሞ	1060	10	070 -	1080 CTT ATA	G11 171	1090
20							
	1.1	100	1110		1120	11	30
	TGT GTG	CCT GTG	AGG TCC	AAG CTA	GAG GAA	GTC CAA	AGA AAA
25					***		
					160		
					TCG GTG		AAT TAT
30		•					
	1180	13	190	1200	·	1210	CTA ATT
	TCC TCT	GAG. TGG	GVG C10			GAT GIT	CIN ATI
35	1220	1220		1240	7	250	1260
_					TGG GCT		
					****	***	
							1300
40			CCT TTT				AGG GAG
	_						
		310	1320	CNA GGI	1330	21	
45	GAA ATT		101 00			5.	

dans laquelle le signe "--" représente une base complémentaire de la base représentée juste au-dessus de chaque signe.

- 3. Vecteur d'expression dans lequel un fragment d'ADN contenant une séquence de bases telle que définie à la revendication 1 ou 2 et codant pour un antigène rencontré spécifiquement chez un hôte atteint de l'hépatite non-A non-B, est introduit dans un site de clonage présent en avai d'un promoteur dudit vecteur.
- 4. Vecteur d'expression selon la revendication 3, dans lequel le promoteur peut être contrôlé par un facteur de régulation.
  - 5. Vecteur d'expression selon la revendication 3, dans lequel le promoteur opère dans un microorganisme.
  - 6. Vecteur d'expression selon la revendication 3, dans lequel le promoteur opère dans un eucaryote.
- 7. Transformant obtenu par transformation d'un hôte par un vecteur d'expression dans lequel un fragment d'ADN contenant une séquence de bases telle que définie à la revendication 1 ou 2 et codant pour un antigène

rencontré spécifiquement chez un hôte atteint de l'hépatite non-A non-B, est introduit dans un site de clonage présent en avail d'un promoteur dudit vecteur.

- 8. Transformant selon la revendication 7, dans lequel l'hôte est Escherichia coli ou Bacillus subtilis.
- 9. Procédé de production d'un antigène rencontré spécifiquement chez un hôte atteint de l'hépatite non-A non-B, comprenant l'introduction d'un fragment d'ADN contenant une séquence de bases telle que définie à la revendication 1 ou 2 et codant pour ledit antigène rencontré spécifiquement dans un site de clonage présent en aval d'un promoteur d'un vecteur d'expression, l'introduction du vecteur d'expression contenant ledit fragment d'ADN dans un hôte, la culture dudit transformant, et la récupération de l'antigène produit et accumulé.
- 10. Procédé pour le diagnostic in vitro de l'hépatite non-A non-B, qui comprend la mise en contact d'un échantilion de foie et/ou d'un échantilion de sérum prélevé chez un patient soupçonné être infecté par une hépatite non-A non-B, avec la protéine dont la séquence apparaît à la revendication 1, ou une partie de celle-ci, pendant un laps de temps suffisant et dans des conditions suffisantes pour permettre la production d'un complexe entre ladite protéine ou une partie de celle-ci, avec les anticorps contenus dans l'échantillon du patient, et la détection de la présence du complexe immunologique, en particulier, lorsque le patient souffre d'une hépatite non-A non-B.

#### Patentansprüche

1. DNA-Fragment, das eine für ein antigenes Protein, das spezifisch in einem mit non-A non-B Hepatitis befallenen Wirt vorkommt, kodierende Basensequenz enthält, wobei das Protein die ganze oder einen Teil der durch die Formel dargestellten Aminosäuresequenz enthält:

55

										10					
	Het	λla	Val	Thr	The	yrd	Leu	Thr	Trp		His	Ğlu	Lys	Ile	Leu
5	Gln	Asn	His	Phe	20 Gly	Gly	Lys	λrg	Leu	Ser	Leu	Leu	Tyr	Lys	30 Gly
10	Ser	Val	His	Gly	Phe	His	Asn	Gly	Val	40 Lau	Lau	Asp	lrg	Cys	Сла
	Asn	Gln	Gly	Pro	50 Thr	Leu	Thr	Val	Ila	Tyr	Ser	Glu	λsp	His	60 Ile
15	Ile	Gly	٨La	Tyr	Ala	Glu	Glu	Gly	Tyr	70 Gla	Glu	Arg	Lys	Tyr	λla
20	Ser	Ile	Ila	Leu	80 Phe	Ala	Leu	Gla	Glu	Thr	Lys	Ile	Ser	Glu	. 90 Trp
	Lys	Leu	Gly	Leu	Tyr	Thr	Pro	Glu	The	100 Leu		Cys	Cys	λsp	Val
25	Ala	Lys	ΪÿΞ	Asn	110 Ser	Pro	The	Asn	Phe	Gla	Ile	. Asy	Gly	· Arg	120   Asn
	Arg	Lys	. Val	. Ile	. Met	Asp	Leu	Lys	Thi	130 : <b>Xe</b> t		ı Ası	ı Lev	. Gly	, Leu
30	Ala	Gli	ı Ast	Cys	140 Thr		Ser	: Ile	s Gli	a Asp	, Ty	r Glı	ı Val	L Phe	150 a Arg
35	Суз	Gli	ı Ası	) Sei	. Leu	. Asp	, G11	ı Arç	g Ly:	160 s Ila	_	s Gl	y Val	l Il	a Glu
	Lev	ı Arş	g Lys	s Sez	170 Leu		ı Səl	- Al	a Le	u Ar	g Th	r Ty	r Gl	u Pr	180 Tyr
40	Gly	r Sez	c Lev	ı Val	l Glr	ı Glı	ı Ile	a Ar	g Il	19 a Le	-	u Le	u Gl	y Pr	o Ile

	CGLY	λla	Gly	Lys	200 Ser	Ser	Phe	Phe	Asn	Ser	Val	λrg	Ser	Val	210 Phe
5	(Gln	Gly	His	Val	Thr	His	Gln	Ala	Leu	220 Val	Gly	Thr	Asn	Thr	Thr
10	età'	Ile	Ser	Glu	230 Lys	Tyr	Arg	Thr	Tyr	Ser	Ile	Хгд	Asp	Gly	240 Lys
	λsp	Gly	Lys	Tyr	Leu	Pro	Phe	Ile	Leu	250 Cys	Asp	Ser	Leu	Gly	Leu
15	Ser	Glu	Lys	Glu	260 Gly	Gly	Leu	Cys	Met	Asp	λsp	Ile	Ser	Tyr	270 Ile
	Leu	Asn	Gly	Asn	Ile	Arg	Asp	Arg	īyī	280 Gln	Phe	Asn	Pro	Met	Glu
20	Ser	Ile	Lys	Leu	290 Asn	His	His	λsp	Tyr	Ile	Asp	Ser	Pro	Ser	300 Leu
25	Lys	λsp	Arg	Ile	His	Суs	Val	. Ala	Phe	310 Val		. Asp	Ala	Ser	Ser
	Ile	Glu	Tyr	Phe	320 Sez		- Gla	. Het	Ile	Val	. Lys	: Ile	LY3	. Arg	330 ; Ile
30	Arg	λrg	, Glu	Lev	. Val	. Asr	n Ala	ı Gly	, Väl	340 L Val		. Va.	L Ala	ı Lev	ı Leu
35	Thr	His	. Val	. Asp	350 Sez		t Asp	) Lev	ı Ile	a Thu	r Ly:	5 G1	y <b>X</b> SI	p Lei	360 u Ile
35	Glu	Ile	e Glu	ı Arg	g Cys	s Vai	l Pro	o Vai	l Ar	370 g Se:		s Le	u Gl	u Gl	u Val
40	Gln	. <b>Ar</b> g	y Lys	Lev	380 1 Gl		e Al	a Le	u Se	r Asj	p Il	e Se	r Ya	l Va	390 1 Sea
	λsn	Tyr	: Ser	Sez	Glu	Trp	Glu	ı Lev	Ast	400 Pro	Val	Lvs	: Ast	Val	l Leu
45					410	ı									420 Phe
50	•									430	)				ı Glu
					440	ı	Gly							,	

<sup>2.</sup> DNA-Fragment nach Anspruch 1, bei dem die Basensequenz die ganze oder einen Teil der durch die Formel dargestellten Phasensequenz enthält:

				1	LO			20			30			4	0	
	5'	ATG	GCA	GTG	λCλ	ACT	CGT	TTG	ACA	TGG	TTG	ĊAT	GAA	λAG	ATC	
5	3'															
3				50			60			•				80		
		CTG	CAA	AAT	CAT	ተጥተ	GGA	GGG	AAG	ccc		) CC	CTT	CTC	ጥልጥ	
10			90			10	00			110			120			
		λAG	GGT	AGT	GTC	CAT	GGA	TTC	CAT	AAT	GGA	GTT	TTG			
		1:	30			140			150			1	60			
15							GGG							TAT	AGT	
		170														
20		Gλλ					GGA									
															~~~	
				2	20			230			240			2	50	
		GAA	AGA	AAG	TAT	GCT	TCC	ATC	ATC	CTT	TIT	GCA	CTT	CAA	GAG	
25																
				260	•		270			2	80			290		
		ACT					TGG									
30															•	
			300			3	10									
		ACA				TGT	GAC	GIT	GCA	. AAA	TAT	AAC	TCC	CCA	ACT	)

	3	40			350			360			3.	70		
	AAT	TTC	CAG	λTA	GAT	GGA	λGA	AAT	λGA	λΑλ	GTG	ATT	ATG	GAC
5	380						40	00			110			420
	TTA	AAG	ACA	ATG	GAA	AAT	CTT	GGA	CTT	GCT	CAA	AAT	TGT	ACT
10			4.	30		4	140			450			41	50
	ATC	TCT	ATT	CAG	GAT	TAT	GAA	GTT	TTT	CGA	TGC	GAA	GAT	TCA
						•								
		4	170			480			49	0		5	00	
15	CTG	GAC	GAA	λGλ	AAG	ATA	AAA	GGG	GTC	ATT	GAG	CTC	AGG	AAG
		510			52	20			30			540		
20	AGC	TTA	CTG	TCT	GCC	TTG	AGA	ACT	TAT	GAA	CCA	TAT	GGA	TCC
		50		•	560		•	570			58	30		
	CTG	GTT	CAA	CAA	λTA	CGA	ATT	CTG	CTG	CTG	CCT	CCA	ATT	GGA
25														
	590			600			63	LO		(	520			630
	GCT	GGG	AAG	TCT	λGC	TTT	TTC	AAC	TCA	GTG	AGG	TCT	CTT	TTC
30														
•			64	10		(	550			660			67	70
	CAA	GGG	CAT	GTA	ACG	CAT	CAG	GCT	TTG	GTG	GGC	ACT	117	161
35			80			690			7(	00		•	710	•
	ACT	GGG	ATA	TCT	GAG	AAG	TAT	AGG	ACA	TAC	TCT	ATT	AGA	GAC
									•					
40		720			7:	30			740			750		
₩.	GGG	AAA	GAT	GGC	AAA	TAC	CTG	CCA	TTT	ATT	CIG	TGT	GAC	TCA
	70	60		•	770			780			79	90		•
45	CTG	GGG	CTG	AGT	GAG	AAA	GAA	GGC	GGC	CTG	TGC	ATG	GAT	GAC
					*									:
	800			810				20		1	830			840
	ATA	TCC	TAC	ATC	TTG	AAC	GGT	AAC	ATT	CGT	GAT	AGA	TAC	CAG
<b>50</b>														
			85	50			860			870			8	80
	TTT	AAT	CCC	ATG	GAA	TCA	ATC	AAA	TTA	AAT	CAT	CAT	GAC	TAC
55														

		1	890			900			91	.0		9	20	
	ATT	GAT	TCC	CCA	TCG	CTG	AAG	GAC	λGλ	ATT	CAT	TGT	GTG	GCA
5														
		020			_									
	ጥጥጥ	GTA	ىك ئىمى	C \ T		10	TC TO	9	50			960		
			***	GVI	500	AGC	ICT	ATT	GAA	TAC	TTC	TCC	TCT	CAC
10														
	91	70		9	980			990			100	00		
	ATG	ATA	GTA	AAG	ATC	AAA	AGA	ATT	CGA	λGG	GAG	TTG	GTA	AAC
15														
10	1010	ccm	CMC	rozo	~~~	cmc	103	30		10	140		1	1050
	GCI	661	GIG	GTA	CAT	GTG	GCT	TTG	CIC	ACT	CAT	GTG	GAT	AGC
	-													
			106	50		10	70		1	080			109	0
20	ATG	GAT	CTG	ATT	ACA	AAA	GGT	GAC	CTT	ATA	GAA	ATA	GAG	AGA
	mc=		100	c m.c	100	1110			112	20		11	130	
25		910		GTG	AGG	TCC	AAG	CTA	GAG	GAX	GTC	CAA	AGA	AAA
		1140			11:	50		11	160		•	1170		
	CTT	GGA	TTT	GCT	CTT	TCT	GAC	ATC	TCG	GTG	GTT	AGC	AAT	TAT
30														
	•													
	mee.	180	C16	13	190			1200			12	10		
	100	101	GAG		GAG	CTG	GAC	CCT	GTA	AAG				
35														
	1220		1	1230			12	40		1:	250			1260
	CII	TCT	GCT	CTG	AGA	CGA	ATG	CTA	TGG	GCT	GCA	GAT	GAC	TTC
														~
40								•	•					
~		636	127	70	^~	13	280	~		1290			13	00
	TTA	GAG				TTT						CTA	λGG	GAG
		12	310		:	1320			13:	30				
45						GCA					3.			

wobei "-" jeweils die zur direkt darüberstehenden Base komplementäre Base darstellt.

- 3. Expressionsvektor, bei dem ein DNA-Fragment, das eine Basensequenz nach Anspruch 1 oder 2 enthält und für ein Antigen kodiert, das spezifisch in einem mit non-A non-B Hepatitis befallenen Wirt auftritt, in eine stromabwärts vom Promotor des Vektors vorhandene Klonierungsstelle eingeführt wird.
  - 4. Expressionsvektor nach Anspruch 3, bei dem der Promotor durch einen Regulationsfaktor kontrollierbar ist.
  - 5. Expressionsvektor nach Anspruch 3, bei dem der Promotor in einem Mikroorganismus wirksam ist.
    - 6. Expressionsvektor nach Anspruch 3, bei dem der Promotor in einem Eukaryonten wirksam ist.
  - 7. Transformante, erhalten durch Transformieren eines Wirtes mit einem Expressionsvektor, bei dem ein DNA-Fragment, das eine Basensequenz nach Anspruch 1 oder 2 enthält, die für ein Antigen kodiert, das spe-

zifisch in einem mit non-A non-B Hepatitis befallenen Wirt auftritt, in eine stromabwärts vom Promotor des Vektors vorhandene Klonierungsstelle eingeführt wird.

8. Transformante nach Anspruch 7, bei der der Wirt Escherichia coli oder Bacillus subtilis ist.

10

20

25

35

*5*0

- 9. Verfahren zur Herstellung eines Antigens, das spezifisch in einem mit non-A non-B Hepatitis befallenen Wirt vorkommt, umfassend die Schritte: Einführen eines DNA-Fragments, das eine Basensequenz nach Anspruch 1 oder 2 enthält, die für das spezifisch vorkommende Antigen kodiert, in eine stromabwärts vom Promotor eines Expressionsvektors vorhandene Klonlerungsstelle, Einführen des das DNA-Fragment enthaltenden Expressionsvektors in einen Wirt, Kultivieren der Transformanten und Gewinnung des produzierten und akkumulierten Antigens.
- 10. Verfahren zur In vitro Diagnose von non-A non-B Hepatitis, umfassend die Schritte: in Kontakt bringen einer Leberprobe und/oder einer Serumprobe, die von einem Patienten mit Verdacht auf eine non-A non-B Hepatitisinfektion entnommen wurde, mit dem Protein mit einer Sequenz wie in Anspruch 1, oder einem Teil davon, für eine bestimmte Zeit und unter Bedingungen, die ausreichend sind für die Komplexbildung zwischen dem Protein, oder einem Teil davon, mit dem in der Patientenprobe enthaltenen Antikörpern, und Nachweis der Anwesenheit des immunologischen Komplexes, insbesondere wenn der Patient an non-A non-B Hepatitis leidet.

### Fig. la

5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG 40 50 60 CAT GAA AAG ATC CTG CAA AAT CAT TTT GGA 70 80 90
GGG AAG CGG CTT AGC CTT CTC TAT AAG GGT 70 100 110 AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG 130 140 CTT GAC AGA TGT TGT AAT CAA GGG CCT ACT 160 170 CTA ACA GTG ATT TAT AGT GAA GAT CAT ATT 190 200 ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG 220 230 GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT 250 260 270 GCA CTT CAA GAG ACT AAA ATT TCA GAA TGG 290 280 AAA CTA GGA CTA TAT ACA CCA GAA ACA CTG

# Fig. 1b

TT	TG:	r rgi	310 <u>C GA</u>	GTI	. GCI	320 <u>A AA</u>	<u> TAI</u>	<u>AA</u>	330 <u>TC</u>
CCA	ACT	3 2 <u>44</u> 2	140 TTC	CAG	ata	350 <u>Ga</u> 1	GGA	<u>AGA</u>	360 <u>AAT</u>
<u>Aga</u>	AA	3 9 <u>79</u>	70 <u>ATI</u>	<u>A</u> TG	GAC	380 <b>TTA</b>	AAG	<u>ACA</u>	390 <u>ATG</u>
<u>GAA</u>	<u>aat</u>	4 CTT	00 <u>GGA</u>	CIT	GCI	410 <u>CAA</u>	<u>aat</u>	<u>TGT</u>	420 <u>ACT</u>
<u>ATC</u>	<u>TCT</u>	4 ATT	30 <u>CAG</u>	<u>Gat</u>	<u>TAT</u>	440 <u>GAA</u>	<u>GTT</u>	TTT	450 <u>CGA</u>
<u> TGC</u>	<u>GAA</u>	4 <u>GAT</u>	60 <u>TCA</u>	<u>CTG</u>	GAC	470 <u>Gaa</u>	<u>aga</u>	<u>aag</u>	480 <u>ATA</u>
<u> </u>	ତ୍ରତ୍ର	4: GIC	90 <u>ATT</u>	<u>GAG</u>	<u>CTC</u>	500 <u>AGG</u>	<u>aa</u> g	<u>agc</u>	510 <u>TTA</u>
<u>Ctg</u>	<u>TCT</u>	5; <u>GCC</u>	20 TTG	<u>aga</u>	<u>act</u>	530 <u>TAT</u>	<u>gaa</u>	CCA	540 <u>TAT</u>
<u>GGA</u>	TCC	55 <u>CIG</u>	O GTT	<u>CAA</u>	CAA	60 <u>ATA</u>	<u>CGA</u>	<u>att</u>	570 <u>CTG</u>
<u>ctg</u>	<u>ctg</u>	58 <u>GGT</u>	0 <u>CCA</u>	<u>ete</u>	GGA	90 <u>GCT</u>	<u> </u>	<u>AAG</u>	600 <u>TCT</u>

# Fig. lc

<u>A</u> GC	TŢŢ	6 TTC	10 <u>AAC</u>	<u>TÇA</u>	<u>GTG</u>	620 <u>A</u> GG	<u> TCT</u>	g <u>t</u> t	630 <u>TTC</u>
<u>CAA</u>	GGG	6 <u>CAT</u>	40 <u>GTA</u>	<u>acg</u>	<u>CAT</u>	650 <u>Cag</u>	<u>GCT</u>	TIG	660 <u>GTG</u>
<u>GGC</u>	act		70 <u>ACA</u>	act	<u> </u>	680 <u>ATA</u>	<u>TCT</u>	GAG	690 <u>AAG</u>
TAT	<u>agg</u>		00 <u>TAC</u>	<u>TCT</u>	<u>att</u>	710 <u>AGA</u>	<u>GAC</u>	GGG	720 <u>AAA</u>
gat	<u>GGC</u>	7. <u>AAA</u>	30 <u>TAC</u>	cig	<u>CCA</u>	740 <u>TTT</u>	att	<u>ct</u> g	750 <u>TGT</u>
gac	<u>TÇA</u>					770 <u>Gag</u>	<u> 777</u>	<u>GAA</u>	780 <u>GGC</u>
<u>G</u> GÇ	CIG		eo <u>Atg</u>	<u>gat</u>	<u>GAC</u>	800 <u>Ata</u>	<u>TCC</u>		810 <u>ATC</u>
<u>TTG</u>	<u>aac</u>		20 <u>AAC</u>	<u>att</u>	CGT	330 <u>GAT</u>	<u>aga</u>		840 <u>CAG</u>
TTT	<u>aat</u>		O <u>ATG</u>	<u>gaa</u>	<u>TCA</u>	B60 ATC	<u> </u>		870 <u>AAT</u>
<u>CAT</u>	<u>Cat</u>	88 <u>GAC</u>	IO TAC	<u>att</u>	GAT	390 <u>TCC</u>	<u>cca</u>	<u>TCG</u>	900 <u>CTG</u>

### Fig. ld

910 920 AMG GAC AGA ATT CAT TGT GTG GCA TTT GTA 950 940 TIT GAT GCC AGC TCT ATT GAA TAC TTC TCC 970 980 . 990 TCT CAG ATG ATA GTA AAG ATC AAA AGA ATT 1000 1010 1020 CGA AGG GAG TTG GTA AAC GCT GGT GTG GTA 1030 1040 1050 CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC 1060 1070 ATG GAT CTG ATT ACA AAA GGT GAC CTT ATA 1090 1100 1110 GAA ATA GAG AGA TGT GTG CCT GTG AGG TCC 1120 1130 1140 AAG CTA GAG GAA GTC CAA AGA AAA CTT GGA 1150 1160 1170 TIT GCT CTT TCT GAC ATC TCG GTG GTT AGC 1180 1190 1200 AAT TAT TCC TCT GAG TGG GAG CTG GAC CCT

# Fig. le

1210 1220 1230
GTA AAG GAT GTT CTA ATT CTT TCT GCT CTG

1240 1250 1260 AGA CGA ATG CTA TGG GCT GCA GAT GAC TTC

1270 1280 1290 TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG

1300 1310 1320
AAT CTA AGG GAG GAA ATT ATC AAC TGT GCA

1330 CAA GGA AAA AAA 3' 5'

Fig. 2

AAAAATTTATTTGCTTTCAGGAAAATTTTTCTGT TTTTTAAATAAACGAAAGTCCTTTTAAAAAAGACA

ATAATGTGTGGAATTGTGAGCGGATAACAATTTC TATTACACACCTTAACACTCGCCTATTGTTAAAG

Fig. 3a

300 CCA Pro	360 GAA G1u	420 TGC Cys	480 CTG Leu	540 CTG Leu
TCC	ATG	CGA	TTA	CTG
AAC	ACA	TTT Phe	AGC	ATT
TAT TYT	350 AAG Lys	410 GTT Val		530 CGA
AAA Lys	TTA Leu	GAA Glu	470 AGG AAG Arg Lys	ATA Ile
GCA Ala	GAC	Tat Tyr	Circ	g g g n
280 GAC GTT ASP VAL	340 ATT ATG Ile Met		GAG Glu	
GAC Asp	34 ATT Ile	400 CAG GAT Gln Asp	460 ATT GAG Ile Glu	520 GTT CAA Val Gln
TGT Cys	GTG Val	ATT Ile	GTC Val	Crc
TGT Cys	AAA Lys	TCT	GGG	FCC
270 TTT Phe	330 AGA Arg	390 ATC Ile	450 AAA Lys	510 GGA Gly
CFG	AAT Asn	ACT	ATA Ile	tat Tyf
ACA	AGA	TGT Cys	AAG Lyb	CCA
260 GAA Glu	320 GGA Gly	380 AAT ABN	440 AGA Arg	500 GAA Glu
S S S	GAT.	450 810 810	Sin Sin	TAT Tyf
ACA Thr	ATA Ile	GCT	GAC	ACT
250 CTA TAT Leu Tyr	310 TTC CAG Phe Gln	CTT	430 TCA CTG Ser Leu	0 Aga Arg
re d	TTC TTC Phe	370 T GGA CTT u Gly Lei	43 TCA Ser	490 TTG AGA Leu Arg
GGA G1y	AAT Asn	Cir	gat Asp	GCC Ala
C. T. P. P. C. T. P. P. C. T. P. P. C. T. P.	ACT	AAT Asn	GAA Glu	TCT

CTG GGT CCA ATT GGA GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC CAA Leu Gly Pro Ile Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe Gln

Fig. 3b

1

660 TAT Tyr	720 Gac Asp	780 TTG Leu	840 CAT His	900 TTT Phe	960 CGA AFG	1020 ATG Met
AAG Lys	TGT Cyb	ATC Ile	AAT Asn	GTA Val	ATT Ile	AGC
GAG Glu	CTG	TAC	TTA Leu	TTT Phe	AGA Arg	GAT Asp
650 ATA TCT Ile Ser	710 ATT Ile	770 ATA TCC Ile Ser	330 AAA Lys	90 GCA Ala	950 ATC AAA Ile Lys	1010 CAT GTG His Val
	TTT Phe	ATA Ile	ATC 11e	GTG Val	ATC Ile	CAT His
GGG G1y	CCA	760 ATG GAT GAC A Met Asp Asp I	TCA Ser	880 ATT CAT TGT GTG GCA Ile His Cys Val Ala	AAG Lys	r F
40 ACT Thr	700 TAC CTG Tyr Leu	50 GAT ASP	820 ATG GAA Met Glu	O CAT His	940 ATA GTA 1 Ile Val 1	1000 Trg crc / Leu Leu ?
6 ACA Thr	7. TAC TYF	76 ATG Met	82 ATG Met	88 ATT Ile	94 ATA Ile	100 116
AAT Asn	AAA Lys	7 g C	CCC	AGA Afg	ATG Met	GCT
ACT	GGC	CTG	AAT Asn	GAC Asp	CAG Gln	GTG
630 GGC GLY	690 Gat Abp	750 GGC G1y	810 TTT Phe	870 AAG Lys	930 TCT Ser	990 CAT His
GTG Val	AAA Lys	GGC G1Y	CAG	CTG Leu	TCC	GTA
TTG	GGG	GAA	TAC	TCG	TTC	GTG
CAG GCT Gln Ala	680 Aga gac Arg asp	740 GAG AAA Glu Lys	800 AGA	860 TCC CCA Ser Pro	920 GAA TAC Glu Tyr	980 GCT GGT Ala Gly
CAG Gln	AGA Arg	GAG Glu	gat Asp	TCC Ser	6 45 u	GCT Ala
CAT	ATT Ile	AGT Ser	CGT	gat Asp	ATT	970 TTG GTA AAC (
610 GTA ACG Val Thr	70 TCT Ser	730 GGG CTG Gly Leu	790 AAC ATT Asn Ile	850 TAC ATT Tyr Ile	TCT Ser	OTA GTA Val
GTA Val	67 TAC Tyr	7. 666 61y	75 AAC Asn	EAC Tyr	910 AGC TCT Ser Ser	970 TTG GI Leu Ve
CAT	ACA	CTG	GGT Gly	GAC	GCC	GAG Glu
666 61y	AGG	TCA	AAC	CAT	GAT	AGG
	•					

Fig. 3c

GAT CTG ATT ACA AAA GGT GAC CTT ATA GAA ATA GAG AGA TGT GTG CCT GTG AGG TCC AAG ASP Leu ile Thr Lys Gly Asp Leu ile Glu arg Cys Val Pro Val Arg Ser Lys 1140 1130 1140 1140 1150 1130 1140 1140 1150 1130 1140 CTA GAG GAA GTC CAA AGA AAA CTT GGA TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC AAT Leu Glu Glu Glu Val Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser Asn TCT GAG TGG GAG CTG GAC CCT GTA AAG GAT GTT CTA ATT CTT TCT GCT CTG AGA Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu Ile Leu Ser Ala Leu Arg 1190 1180 1170 1160 1150 JCC J Tyr Ser TAT

1270 1280 1290 1300 CTA AGG GAG GAA ATT ATC AAC TGT GCA CAA GGA AAA AAA TAG Leu Arg Glu Glu Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*

80	160	240	320	400	480	560
TCGTTTGACA	TCCATGGATT	Catattattg	Taaaatteca	CTAATTTCCA	TGTACTATCT	GCTCAGGAAG
AGCAAACTGT	AGGTACCTAA	Gtataataac	Attttaaagt	GATTAAAGGT	ACATGATAGA	CGAGTCCTTC
70	150	230	310	390	470	550
CAGTGACAAC	AAGGGTAGTG	Tagtgaagat	TTCAAGAGAC	Aactccccaa	TGCTCAAAAT	GGGTCATTGA
GTCACTGTTG	TTCCCATCAC	Atcacttcta	AAGTTCTCTG	Ttgaggggtt	ACGAGTTTTA	CCCAGTAACT
60	140	220	300	380	460	540
AGAAGTATGG	CCTTCTCTAT	Cagtgattta	Ctttttgcac	TGCAAAATAT	Atcttggact	Aagataaaag
TCTTCATACC	GGAAGAGATA	Gtcactaaat	Gaaaaacgtg	ACGTTTTATA	Tagaacctga	Ttctattte
50	130	210	290	370	450	530
Caacagatca	AGCGGCTTAG	CCTACTCTAA	TTCCATCATC	GTTGTGACGT	Acaatggaaa	Ggacgaaaga
Gttgtctagt	TCGCCGAATC	GGATGAGATT	AAGGTAGTAG	CAACACTGCA	Tgttaccttt	Cctgctttct
40	120	200	280	360	440	520
Acagacagta	TTTGGAGGGA	Taatcaaggg	Gaaagtatgc	ACACTGTTT	GGACTTAAAG	Aagaitcact
TgTCTGTCAT	AAACCTCCCT	Attagttccc	Ctttcatacg	TGTGACAAAA	CCTGAATTTC	Tictaagiga
30	110	190	270	350	430	510
AGCTCATACT	GCAAAATCAT	ACAGATGTTG	Taccaggaaa	Tacaccagaa	Aagtgattat	TTTCGATGCG
TCGAGTATGA	CGTTTTAGTA	TGTCTACAAC	Atggtccttt	Atgtggtctt	TTCACTAATA	AAAGCTACGC
20	90	180	260	340	410	500
CCTCAGCTCT	TGGTTGCATG AAAAGATCCT	GTTTTGCTTG ACAGATGTTG	AGAAGAGGGT	TAGGACTATA	GATAGATGGA AGAAATAGAA	Ttatgaagtt
GGAGTCGAGA	ACCAACGTAC TFTTCTAGGA	CAAAACGAAC TGTCTACAAC	TCTTCTCCCA	ATCCTGATAT	CTATCTACCT TCFFFATCTT	Aatacttcaa
10	90	170	250	330	410	1 490
GGGGGCTAC	TGGTTGCATG	CCATAATGGA	GAGCATATGC	GAATGGAAAC TAGGACTATA	GATAGATGGA	CTATTCAGGA
CCCCCCATG	ACCAACGTAC	GGTATTACCT	CTCGTATACG	CTTACCTTTG ATCCTGATAT	CTATCTACCT	GATAAGTCCT
30.00						

640	720	800	880	960	1040	1120
TGGGTCCAAT	TTGGTGGGCA	GCCATTTATT	ACGTAACAT	TCGCTGAAGG	AAAGATCAAA	ATCTGATTAC
ACCCAGGTTA	AACCACCCGT	CGGTAAATAA	TGCCATTGTA	AGCGACTTCC	TTTCTAGTTT	TAGACTAATG
630	710	790	870	950	1030	1110
ATTCTGCTGC	GCATCAGGCT	GCAAATACCT	TACATCTTGA	TGATTCCCCA	AGATGATAGT	GATAGCATGG
TAAGACGACG	CGTAGTCCGA	CGTTTATGGA	ATGTAGAACT	ACTAAGGGGT	TCTACTATCA	CTATCGTACC
610 620	700	770	860	940	1020	1100
CCCTGGTTCA ACAAATACGA	GGCATGTAAC	TATTAGAGAC GGGAAAGATG	TGACATATCC	Atgactacat	TTCTCCTCTC	CACTCATGTG
GGGACCAAGT TGTTTATGCT	CCGTACATTG	ATAATCTCTG CCCTTTCTAC	ACTGTATAGG	Tactgatgta	AAGAGGAGAG	GTGAGTACAC
610	690	770	850	930	1010	1090
CCCTGGTTCA	GTTTTCCAAG	TATTAGAGAC GGGAAAGATG	TGTGCATGGA	TTAAATCATC	Tattgaatac	TGGCTTTGCT CACTCATGTG
GGGACCAAGT	CAAAAGGTTC	ATAATCTCTG CCCTTTCTAC	ACACGTACCT	AATTTAGTAG	Ataacttatg	ACCGAAACGA GTGAGTACAC
600	680	760	840	920	1000	1080
CCATATGGAT	AGTGAGGTCT	GGACATACTC	GAAGGCGCC	ATCAATCAAA	Atgccagctc	GTGGTACATG
GGTATACCTA	TCACTCCAGA	CCTGTATGAG	CTTCCGCCGG	TAGTTAGTTT	Tacggtcgag	CACCATGTAC
590	670	750	830	910	990	1070
AACTTATGAA	TITICAACIC	GAGAAGTATA	Gagtgagaaa	ATCCCATGGA	Titgiaithg	AAACGCTGGT
TTGAATACTT	AAAAGIIGAG	CTCTTCATAT	Ctcactcttt	TAGGGTACCT	Aacataaac	TTTGCGACCA
570 580	660	740	820	900	980	1050
AGCTTACTGT CTGCCTTGAG	AAGTCTAGCT	TGGGATATCT	CACTGGGGCT	Taccagtita	TTGTGTGGCA	AGAATTCGAA GGGAGTTGGT
TCGAATGACA GACGGAACTC	TTCAGATCGA	ACCCTATAGA	GTGACCCCGA	Atggtcaat	AACACCGT	TCTTAAGCTT CCCTCAACCA
570	650	730	810	890	970	1050
AGCTTACTGT CTGCC	TGGAGCTGGG	CTAATACAAC TGGG	CTGTGTGACT	TCGTGATAGA	ACAGAATTCA	Agaattcgaa
TCGAATGACA GACGC	ACCTCGACCC	GATTATGTTG ACCCT	GACACACTGA	AGCACTATCT	TGTCTTAAGT	Tcptaagctt

Fig. 4b

1200	1280	1360	1440	1520	1600
CTTGGATTTG	AATECTITCT	Taagggagga	Agattaaaat	Gatgaagaaa	Aataatitit
GAACCTAAAC	TTAAGAAAGA	Aiftccttcct	Tctaatttta	Ctacttcttt	Ttattaaaa
1190	1270	1350	1430	1510	1590
CCAAAGAAAA	AGGATGTTCT	Atagggaatc	Acatcacaga	Taatgtctag	Gaaaaataat
GGTTTCTTTT	TCCTACAAGA	Tatcccftag	Tgtagtgtct	Attacagatc	Cttttatta
1180	1260	1340	1420	1500	1580
TAGAGGAAGT	GACCCTGTAA	TTTTGAGCAA ATAGGGAATC	AAATTTCCTC	TGTGTTTTAT	TCATAATTGT
ATCTCCTTCA	CTGGGACATT	AAAACTCGTT TATCCCTTAG	TTTAAAGGAG	ACACAAATA	AGTATTAACA
1160 1190 1190 1190 1190 AGTCCAAGC TAGAGAAGT CCAAAGAAAA ACACGGACAC TCCAGGTTCG ATCTCCTTCA GGTTTCTTTT	1250 1260 1270	1330 1340 1350	1410	1490	1570 1580 1590 1600
	GTGGGAGCTG GACCCTGTAA AGGATGTTCT	AGGATTTGCC TTTTGAGCAA ATAGGGAATC	AGGTTCACGT	Accaaagga	CATGATTTAG TCATAATTGT GAAAAATAAT AATAATTTT
	CACCCTCGAC CTGGGACATT TCCTACAAGA	TCCTAAACGG AAAACTCGTT TATCCCTTAG	TCCAAGTGCA	Tggtttcct	GTACTAAATC AGTATTAACA CTTTTATTA TTATTAAAA
1150 1160 1170 1180 1190 TAGAGAGATG TGTGCCTGTG AGGTCCAAGG TAGAGAAGT CCAAAGAAAA ATCTCTCTAC ACAGGACAC TCCAGGTTCG ATCTCCTTCA GGTTTCTTTT	1240	1320	1400	1480	1560
	Aftectea	GACTTCTTAG	GATATGTGAA AGGTTCACGT	AGTAACTAAG	CTAGAAATAA
	Taagagact	CTGAAGAATC	CTATACACTT TCCAAGTGCA	TCATTGATTC	GATCTTTATT
1150	1230	1310	1390	1470	
TAGAGAGATG	GTTAGCAAIT	GGCTGCAGAT	Gaaaaaaata	ACCAAAGAGA	
ATCTCTCTAC	CAATCGTTAA	CCGACGTCTA	Ctttttttt	TGGTTTCTCT	
1140	1210	1290 1300 1310	1380	1450 1460 1470 1480 1490 1500 1510 TCAGAAAGGA ACCAAAGGGA TGTGTTTTAT TAATGTCTAG AGTCTTTCCT AGTCTTTCCT TGGTTTCCT AGTCTTTCCT TGGTTTCCT ACCAAAATA ATTACAGATC	1530 1540 1550
CTTATAGAAA	CTCTTTCTGA CATCTCGGTG	GCTCTGAGAC GAATGCTATG GGCTGCAGAT	TGTGCACAAG		TGCATAGAAC ATTGTAGTAC TTGTAAATAA
GAATATCTTT	GAGAAAGACT GTAGAGCCAC	CGAGACTCTG CTTACGATAC CCGACGTCTA	ACACGTGTTC		ACGTATCTTG TAACATCATG AACATTTATT
1130	1210	1290	1370	1450	1530
AAAAGGTGAC	CTCTTTCTGA	GCTCTGAGAC	AATTATCAAC	TCAGAAAGGA	TGCATAGAAC
TTTCCACTG	GAGAAGACT	CGAGACTCTG	TTAATAGTTG	AGTCTTTCCT	ACGTATCTTG

Fig. 4c

Fig.5

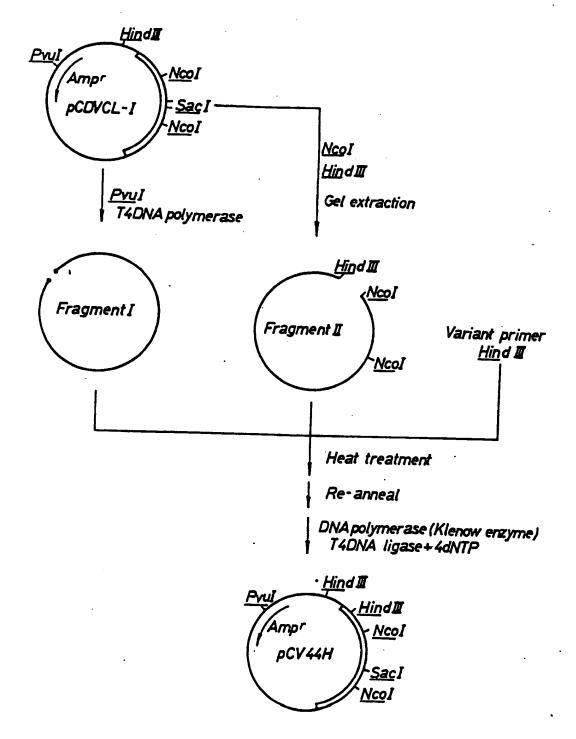


Fig.6

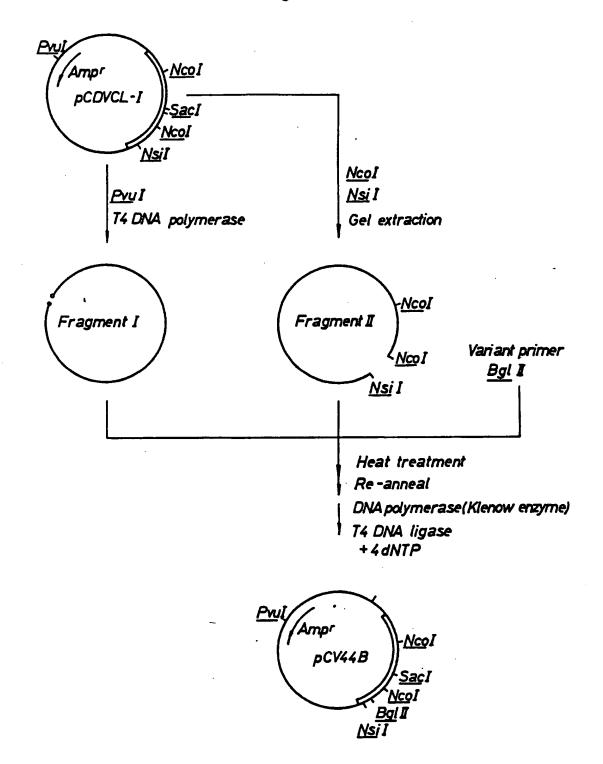


Fig.7

